

Washington University in St. Louis Washington University Open Scholarship

Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Fall 12-18-2017

Exploring the Underlying Mechanisms of Structure Building

Reshma Gouravajhala

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds



Part of the [Cognitive Psychology Commons](#)

Recommended Citation

Gouravajhala, Reshma, "Exploring the Underlying Mechanisms of Structure Building" (2017). *Arts & Sciences Electronic Theses and Dissertations*. 1179.

https://openscholarship.wustl.edu/art_sci_etds/1179

This Thesis is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological & Brain Sciences

Exploring the Underlying Mechanisms of Structure Building
by
Reshma Gouravajhala

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Masters of Arts

December 2017
St. Louis, Missouri

© 2017, Reshma Gouravajhala

Table of Contents

List of Figures	iv
List of Tables	v
Acknowledgments.....	vii
Abstract	viii
Chapter 1: Introduction	1
1.1 What is Structure Building?	2
1.2 How is Structure Building Measured?	4
1.3 Is Structure Building the Same as Reading Comprehension?	4
1.4 Is Structure Building Important in the Classroom?	6
1.5 The Present Study.....	8
Chapter 2: Method	16
2.1 Participants and Design.....	16
2.2 Materials	16
2.3 Procedure.....	19
Chapter 3: Results	21
3.1 Main Point Identification Task.....	26
3.2 Short-Answer Test.....	31
3.3 Relatedness Ratings Task.....	35
Chapter 4: Discussion	46
4.1 Main Point Identification Task.....	46
4.2 Short-Answer Test.....	49
4.3 Relatedness Ratings Task.....	51
4.4 What do Individual Differences in Structure Building Involve?	52
4.5 Educational Implications of Textbook Scaffolding	53
4.6 Conclusion.....	55
References	58
Appendix A.....	61
Appendix B	62

Appendix C	63
Appendix D.....	65

List of Figures

Figure 3.1: Distribution of MMCB Scores.....	21
Figure 3.2: Relationship between Standardized MMCB and Working Memory	22
Figure 3.3: Relationship between Standardized MMCB and Mindwandering	23
Figure 3.4: Relationship between Standardized MMCB and Number of Working Memory Intrusions	23
Figure 3.5: The Interaction between Text Condition and Standardized MMCB on Short-answer Test Performance Accounting for Working Memory and Mindwandering	32
Figure 3.6: The Interaction between Text Condition and Standardized MMCB on Short-answer Test Performance, Accounting for Main Point Identification Accuracy.....	34
Figure 3.7: The Average Median Expert Network and Examples for the First Relatedness Ratings Task Set	37
Figure A.1: Text Comparison.....	61

List of Tables

Table 3.1: Descriptive Statistics of Predictors and Dependent Variables in Tasks.....	24
Table 3.2: Correlations of Predictors and Dependent Variables in Tasks.....	25
Table 3.3: Results Showing the Main Effects of Text Condition and MMCB on the Standardized Number of Hits in the Main Point Identification Task, Accounting for Working Memory and Mindwandering	27
Table 3.4: Results Showing the Null Effects of All Predictors on the Standardized Number of False Alarms in the Main Point Identification Task	28
Table 3.5: Results Showing the Main Effect of MMCB on d' in the Main Point Identification Task, Accounting for Working Memory and Mindwandering, and Text Condition..	29
Table 3.6: Results Showing the Main Effects of Text Condition and MMCB on a More Liberally Calculated d' in the Main Point Identification Task, Accounting for Working Memory and Mindwandering	30
Table 3.7: Results Showing the Interaction between Text Condition and MMCB on Short-answer Test Performance, Accounting for Working Memory and Mindwandering, and Text Condition	33
Table 3.8: Results Showing the Main Effect of MMCB and the Interaction between Text Condition and MMCB on Short-answer Test Performance, Accounting Main Point Identification Task Accuracy (d') and Text Condition.....	35
Table 3.9: Network Coherence with Imputed Data in the First Relatedness Rating Set	36
Table 3.10: Coherence of Experts' Networks in All Three Relatedness Ratings Tasks	36
Table 3.11: Path-link Correlations of Experts' Networks to Average Median Network in All Three Relatedness Ratings Tasks	37
Table 3.12: Results Showing the Main Effect of MMCB on Standardized Similarity Scores in the First Relatedness Rating Task Set, Accounting for Working Memory and Mindwandering, and Text Condition	38
Table 3.13: Results Showing the Effects of Working Memory and Mindwandering on Standardized Similarity Scores in the First Relatedness Rating Task Set	39
Table 3.14: Results Showing the Main Effect of Text Condition on Standardized Similarity Scores in the Second Relatedness Rating Task Set, Accounting for Working Memory and Mindwandering, and MMCB	40

Table 3.15: Results Showing the Effects of Working Memory and MMCB on Standardized Similarity Scores in the Third Relatedness Rating Task Set, Accounting for Working Memory and Mindwandering, and Text Condition	41
Table 3.16: Results Showing the Effects of Working Memory on Standardized Similarity Scores in the Third Relatedness Rating Task Set, Accounting for Mindwandering	42
Table 3.17: Results Showing the Main Effect of MMCB on Standardized Similarity Scores in the First Relatedness Rating Task Set, Accounting for Main Point Identification Accuracy (d')	43
Table 3.18: Results Showing the Main Effect of Text Condition on Standardized Similarity Scores in the Second Relatedness Rating Task Set, Accounting for MMCB Main Point Identification Accuracy (d')	43
Table 3.19: Results Showing the Main Effects of MMCB and Main Point Identification Accuracy (d') Standardized Similarity Scores in the Third Relatedness Rating Task Set, Accounting for Text Condition	44

Acknowledgments

I would like to express my sincere gratitude to my advisor, Mark McDaniel, for his unfailing support and guidance over the course of my graduate career. I would like to thank my other committee members, Roddy Roediger and Julie Bugg, for their immensely helpful feedback, and for the fun facts about Larry Jacoby's collegiate football career and Pavlov's 14 mixed-breed dogs that now firmly exist in my mental model of the history of psychology.

I would like to thank April Bednarski and Elise Walck-Shannon, who graciously lent their time and Biology expertise, and Siera Stoen, who patiently answered my many questions about Pathfinder and R. I would also like to thank Rebecca Meza, Jack Basse, and Ana Boyer, for their diligence in developing materials, collecting data, and scoring participants' responses.

I am especially thankful to my labmates, Francis Anderson and Toshi Miyatsu, for the brainstorming lunch sessions, the impromptu stats lessons, the numerous pep talks, and the controlled allowance of coffee. I would like to thank Khuyen Nguyen for the dozens of phone calls full of suggestions and support. I would further like to thank my friends, especially Eylül Tekin, Samuel Chung, Marina Gross, and Öykü Uner, for the work parties and *Hamilton* sing-alongs. I am particularly grateful to Eric Failes, for not only his constructive edits about working memory analyses and method section how-tos, but also the unwavering faith in my abilities.

Lastly, I am forever grateful to my parents and my older brother (a true role model!) for their endless support, advice, and encouragement.

Reshma Gouravajhala

Washington University in St. Louis

December 2017

Abstract

Exploring the Underlying Mechanisms of Structure Building

by

Reshma Gouravajhala

Master of Arts in Psychological & Brain Sciences

Washington University in St. Louis, 2017

Professor Mark McDaniel

Structure building, the ability to build a coherent mental model of any narrative, requires the identification and integration of important parts of that narrative, as well as the suppression of irrelevant details. Critically, while individual differences in structure building have been shown to have important consequences in the classroom, little has been concluded about underlying deficits and causal mechanisms of low structure building ability. In the present study, we tested the theory that an impaired ability to suppress unimportant details is low structure builders' sole deficit (Gernsbacher, 1990). We presented participants with educationally authentic text materials that offered varying degrees of structural support, and tested whether structure building predicted their performance, after accounting for working memory and mindwandering, on a main point identification task, a short-answer test of deep-level questions, and a relatedness ratings task. Contradicting the existing theory, we found that those with low structure building ability experienced (relative to high structure builders) a previously unknown deficit: an impaired ability to identify the most important parts of the text. We also found structure-building-related performance differences on our two other comprehension measures; notably, these differences could not fully be accounted for by the main point identification deficit. Lastly,

we affirmed current textbook scaffolding practices, but also identified areas needing further improvement in order to specifically bolster low structure builders' abilities.

Chapter 1: Introduction

In order to succeed, it is imperative that college students are able to organize the information they learn – from classroom lectures, textbooks, or discussions – into foundational main points without getting lost in superficial details. In other words, it is crucial that students do not lose the forest for the trees.

This ability, referred to as *structure building*, is derived by one's success in creating a "cohesive, mental representation or 'structure'" of any event (Gernsbacher, Varner, & Faust, 1990). At the broad level, successful structure building requires the identification and integration of main points, as well as the suppression of irrelevant details. Take, for example, a topic like classical conditioning, which is a staple in introductory psychology courses: students must be able to understand what concepts such as unconditioned stimuli and conditioned responses mean individually and as part of the overarching process, without focusing on how many times Pavlov rang his bell or what dog breeds he conducted research on. This prioritization strategy helps create a strong foundation of important points, and aids in the elaboration of a more stable network of details. The ability to create these hierarchical structures is thought to strengthen conceptual understanding, and lead to increased comprehension.

Importantly, there exist individual differences in people's structure building abilities. The present study seeks to better understand the cognitive mechanisms involved in structure building by identifying where exactly in the process those with low structure building ability falter. Specifically, are these low structure builders unable to even identify the most important parts of a narrative, or does their deficit lie primarily in the integration of main points? Moreover, the study

could inform potential future interventions designed to bolster low structure builders' abilities. While there have been some efforts to create such interventions (see Arnold, Daniel, Jensen, McDaniel, & Marsh, 2016; Bui & McDaniel, 2015; and Callender & McDaniel, 2007), a more concrete understanding of the underlying mechanisms of structure building is critical for the further development of successful interventions.

1.1 What is Structure Building?

In her chapter describing the structure building framework, Gernsbacher (1990) outlined three structure building processes that underlie various language comprehension phenomena: laying a foundation, mapping incoming information onto developing structures, and shifting to create new structures, if necessary.

The first process is the initial activation of memory nodes, which are the building blocks of mental models. Some evidence for such a process comes from research indicating that individuals spent more time reading the first sentence of a story episode (Haberlandt, 1984) and were more likely to recall a sentence when cued by its first content word (Bock & Irwin, 1980). Interestingly, the latter effect did not emerge if the first sentence or content word were not conducive to building a comprehensive structure around them (Foss & Lynch, 1969). In other words, it appears that this process is dependent on the activation of *important* memory nodes that are tied to topic sentences or content words.

The second process in structure building leads to the expansion of a developing structure by mapping coherent incoming information onto existing structures. The better the new information coheres to the present structure, the more likely it is to activate relevant memory nodes. If incoming information does not cohere to the current foundational memory nodes, however, it may activate different memory nodes, resulting in the formation of a new

substructure. It follows that sentences that matched the conceptual or syntactic structure of previous sentences should be read at a faster pace than those that do not match (Gernsbacher & Robertson, 1994). By mapping related information onto existing foundations, people are better able to build coherent mental models that will later lead to better processing and understanding of a narrative.

The final process further clarifies what happens when important incoming information does not cohere to an already existing structure. Gernsbacher refers to this process as *shifting*, the creation of new substructures that represent new events or episodes that arise within the narrative. Therefore, most mental representations comprise many branching substructures. To this end, individuals took longer to comprehend words or sentences denoting changes in physical or temporal setting (Anderson, Garrod, & Sanford, 1983).

Gernsbacher (1990) also outlined the roles of two mechanisms – enhancement and suppression – in the modulation of the memory nodes’ levels of activation. Enhancement occurs when the information represented by a memory node is essential for future structure building, whereas suppression happens if such information is no longer as necessary. According to Gernsbacher, anaphoric devices (e.g., pronouns, repeated noun phrases, etc.) are integral in this process because they improve mental accessibility of their referents, and highlight the importance of certain concepts over others (i.e., those with fewer or less explicit anaphora) as individuals build structures. In order to build a cohesive mental representation, individuals must actively suppress irrelevant information that does not directly lead to comprehension of the narrative. It is here that Gernsbacher believes low structure builders falter: she asserts that they are unable to dampen the activation of inappropriate information, which negatively impacts the ongoing structure building process. Studies showing that less skilled comprehenders were less

efficient at suppressing irrelevant meanings of ambiguous words, typical-but-absent objects in scenes, word labels on pictures, or incorrect forms of homophones offer support for this idea (Gernsbacher, 1990; Gernsbacher, 1994; Gernsbacher & Faust, 1991). As such, it could be the case that low structure builders' enhancement mechanisms are fully functional, but they still falter in the ability because of faulty suppression mechanisms.

1.2 How is Structure Building Measured?

Gernsbacher designed the Multi-media Comprehension Battery (MMCB; Gernsbacher & Varner, 1988), which consists of multiple stories that are each presented in three modalities: written, auditory, or pictorial. A series of multiple-choice comprehension questions follows each story presentation, and individuals' scores on these questions reveal their structure building ability. It has been found that comprehension scores from the different modalities are highly correlated (with r s ranging from 0.72 to 0.92), and that all three appear to uncover the same underlying general comprehension ability (Gernsbacher et al., 1990). Because of the different modalities' high correlation, it is common practice (see Arnold et al., 2016 and Callender & McDaniel, 2007) to use just the written portion of the MMCB to measure individuals' structure building ability. Past research has also shown the MMCB to be highly reliable ($\alpha = .99$; Gernsbacher et al., 1990).

1.3 Is Structure Building the Same as Reading Comprehension?

Previous research has assessed MMCB's theoretical association with reading abilities indexed by standard reading tests such as the Nelson-Denny Reading Test (NDRT). At the surface level, it may appear that structure building and reading comprehension tap into the same ability – and indeed, Maki, Jonas, and Kallod (1994) found a significant correlation ($r = .46$) between the two

measures – but McDaniel, Hines, and Guynn (2002) showed that, in fact, low structure builders exhibit different deficiencies than do low NDRT readers. In that study, participants were presented with sentences from folk tales in either a scrambled order (and were tasked to rearrange them for coherence) or sentences in an already intact form. Low and high NDRT readers performed equivalently in both the scrambled or intact conditions. By contrast, low structure builders fared worse than high structure builders in the intact condition, but improved their performance to be in line with high structure builders in the scrambled condition. Presumably, the task forced low structure builders to engage in the sort of organizational processing that may come naturally to high structure builders. On the other hand, low NDRT readers' deficit does not lie in organizational processing, and thus they did not gain additional benefits when performing the unscrambling task.

Lastly, there exist a few other administrative distinctions between the two measures: unlike readers taking the NDRT, those taking the MMCB are allowed as much time as they require when reading the passages, but are not allowed access to the texts when answering the questions. Moreover, past research has shown that there are no significant differences in the amount of time low and high structure builders spend reading educationally relevant texts, which is often assumed to be low NDRT readers' underlying deficiency (Martin, Nguyen, & McDaniel, 2016). Therefore, while both the MMCB and the NDRT predict reading comprehension, the former seems to focus on the higher-level processes involved in building mental models, while the latter seems to index the speed of word-level processing (McDaniel et al., 2002). As such, the MMCB appears to explicitly assess the quality of a reader's mental model of a narrative.

1.4 Is Structure Building Important in the Classroom?

Because Gernsbacher's (1990) framework was developed specifically for narrative comprehension, there has been limited research investigating the role of individual differences in structure building ability when learning educationally authentic materials (either in the laboratory or in classroom settings). However, the preliminary evidence is promising. Two studies have examined whether structure building predicts academic success in actual classrooms. Maki and Maki (2002) explored whether course format (in-person or online) would impact learning in an introductory psychology course, and the MMCB was used to infer comprehension skill. Students with high MMCB scores benefitted the most from the online course format, whereas those with low MMCB scores accrued no benefit. Furthermore, MMCB predicted exam performance for students in both course formats. Additionally, for students in the web course, MMCB predicted performance on a post-course sample Psychology GRE exam questions.

Arnold et al. (2016) extended this work by looking at the relationship between MMCB and course grades in college introductory psychology and biology courses. In order to evaluate this relationship after controlling for standard predictors, researchers collected high school GPAs and SAT scores (from a subset of students in the psychology course), as well as Biology Concept Inventory (BC) and Lawson's Classroom Test for Scientific Reasoning (LCTSR) scores (from a subset of students in the biology course). Higher MMCB scores were associated with better high school GPAs and higher SAT verbal scores, but not with SAT math or either biology knowledge measure. Students with greater structure building ability performed better than students with lower structure building ability in both the psychology and biology courses, and critically, it was found that differences in structure building ability predicted final grades in both courses even

after taking the other standard predictors into account. Therefore, the results of these two studies suggest that structure building is a general comprehension skill that appears to predict success in multiple subject domains.

Several studies have also examined the relationship between structure building and learning using authentic classroom-type material in the laboratory. For example, Callender and McDaniel (2007) tested whether individual differences in structure building would impact learning of a social psychology textbook chapter, and found that, overall, low structure builders performed worse on tests of comprehension and memory than high structure builders. Importantly, the authors found that low structure builders benefitted from embedded questions (a study adjunct aimed at promoting a more coherent representation of the text) whereas high structure builders did not. In this same vein, Bui and McDaniel (2015) suggested that providing aids (in the form of illustrated diagrams) during lectures can help scaffold the building of a mental model, and thus found that these aids improved low structure builders' free recall and problem-solving performance.

Lastly, Martin et al. (2016) used two scientific passages on brakes and pumps, and focused on identifying the different levels (propositional, text-based, or situational model) of mental representation at which low structure builders might struggle. Students' performance on recitation, free recall after restudy, and factual multiple-choice tests was assessed. Low structure builders performed significantly worse than high structure builders on all three tests, perhaps reflecting their impaired representational structures. Furthermore, although they did not differ from high structure builders in their initial study time, low structure builders did falter in their metacognitive regulation of restudy time. More specifically, their lower final recall performance

indicated that low structure builders struggled to successfully focus their restudy on information they judged to be not well learned.

Together, these results reflect the importance of structure building in learning and comprehension of didactic materials, as well as a significant predictor of college performance. Some of this work also indicates the high potential for interventions specifically aimed at helping low structure builders. It is important to further elucidate the cognitive mechanisms by which low structure builders create mental models, with the hope that, by identifying where exactly in the structure building process low structure builders falter, it will be easier to design and implement scalable and low-demand interventions in the classroom.

1.5 The Present Study

According to Gernsbacher (1990), low structure builders struggle because of faulty suppression mechanisms; in other words, they are less able to dampen the activation of incoming irrelevant information. This inhibition deficit leads low structure builders to shift too often and thus create cluttered structures, which has negative cascading effects on their mental models. Critically, however, it is possible that low structure builders might exhibit shortcomings in other parts of the structure building process as well, which would also lead to worse performance on later comprehension measures relative to high structure builders.

There are several theoretical implications of poor structure building ability: (a) low structure builders struggle to activate certain memory nodes because they fail to identify the most important concepts, leading to incomplete structures; (b) low structure builders build many substructures around a variety of points that they deem to be important, leading to cluttered structures; (c) low structure builders are less successful in integrating knowledge across parts of their mental models; and (d) low structure builders' organization of main terms is less similar to

experts' representations. Without knowing exactly where in the structure building process low structure builders falter, it is harder to pinpoint how their mental models differ from those built by high structure builders, and how these differences may help explain variances in later task performance.

To this end, we presented undergraduate students with an educationally relevant textbook chapter, and tested their learning and memory of the material two days later. In order to identify where in the structure building process (identifying main points, integrating these points, or both) low structure builders fall behind, we designed specific assessments aimed at tackling each step.

We created a main point identification task, where participants were asked to pick out what they believed to be the most important points from a list of both important and unimportant concepts. We tested participants' ability to combine knowledge from multiple sections of the textbook chapter by giving them a short-answer test of conceptual questions. Lastly, following Callender and McDaniel (2007), we asked participants to provide relatedness ratings for important terms from the text in order to externalize their mental representations of the material. We hoped this would lead to a more direct understanding of the organizational differences (if any) in the mental models built by participants of differing structure building ability. Using a correlation-like index adapted from Goldsmith, Johnson, and Acton (1991), we compared participants' ratings and the resulting structures to the two experts' ratings and structures to assess the similarity between participants' networks and the averaged expert network. Similarity is the number of links shared by two networks (i.e., each participant's network and the average median expert network) over the number of links found in either of the two. Britton and Gulgöz (1991) have shown this measure to be sensitive to readers' ability to create structures that better

matched experts' representations after reading a repaired text as opposed to an impaired text, but it is unclear if this sensitivity expands to differences in structure building ability.

We made several predictions about performance differences on all these tasks between those with low and high structure building ability. On the main point identification task, we predicted that, if Gernsbacher's (1990) faulty suppression mechanism account of structure building is accurate, then those with lower structure building ability would identify more unimportant points than their high structure building counterparts. Further, if individual differences in structure building can be entirely explained by an impaired ability to suppress irrelevant details, there should be no differences in the identification of important points between low and high structure builders. However, we theorized that those with lower structure building ability may struggle at an even earlier point in the structure building process: identifying the main concepts that serve as foundations for their networks. This deficit could also help account for previous findings in the literature, but has yet to be explored. We thus predicted that low structure builders would identify fewer main concepts (as identified by two Biology professor experts) than high structure builders.

We predicted that low structure builders would perform worse on the short-answer test than high structure builders (as was the case in Martin et al., 2016, where low structure builders fared worse on both inference multiple-choice and problem-solving problems), though this could be for multiple reasons. If low structure builders' deficits can entirely be explained by a faulty suppression mechanism, as Gernsbacher (1990) asserts, then they would struggle to answer short-answer questions because of their cluttered mental models. Low structure builders could also perform poorly on this task (relative to high structure builders), which requires the integration of knowledge across many foundational nodes, if they fail to even identify those

important concepts. Therefore, we believed that performance on the main point identification task would subsequently predict performance on the short-answer test.

Lastly, we predicted that low structure builders' networks would show less similarity (than those built by high structure builders) to experts' networks. This would be in accordance with both Gernsbacher's (1990) faulty suppression mechanism theory, as well as the alternative possibility that low structure builders struggle with the even more fundamental issue of identifying main points. Interestingly, Callender and McDaniel (2007) found that structure building did not significantly predict coherence (a measure of internal consistency and network stability) of participants' networks, indicating that perhaps those with lower structure building ability do not struggle to be consistent in their relatedness ratings. Since the coherence measure offers no insight into the accuracy of participants' relatedness ratings (as compared to experts'), we chose to instead measure the similarity of participants' networks to the averaged expert network in the present study.

In addition to these comprehension and retention tasks, which will help determine if low structure builders falter at a more basic part of the structure building process than Gernsbacher (1990) theorized, we also included measures of working memory and mindwandering in the present study, for several reasons. Because both these cognitive measures could theoretically impact task performance, we were interested in the relationships between the measures and structure building ability. Given some unpublished research in our lab indicating that working memory (as measured by the Automated Operation Span) and MMCB are not highly correlated ($r = .13$; $n = 264$), we predicted that, even if there is a significant positive relationship between working memory and MMCB in the present study, low structure builders' deficits cannot be explained exclusively by working memory problems. Though the relationship between

mindwandering and MMCB has, to date, never been empirically examined, we predicted a negative relationship between the two constructs. Ultimately, we were interested to see if structure building ability would predict task performance on each of our three comprehension tasks after accounting for working memory and mindwandering.

In a more exploratory vein, we theorized that our measures of working memory and mindwandering could act as proxies for inhibition, as they have both been implicated in individual differences in text comprehension and retention (Agarwal, Finley, Rose, & Roediger, 2017; Daneman & Merikle, 1996; McVay & Kane, 2012; Redick, Heitz, & Engle, 2007). Interestingly, past research has indicated that individual differences in inhibitory processes can lead to differences in working memory capacity (Hasher & Zacks, 1988; Kane, Bleckley, Conway, & Engle, 2001). According to Hasher, Zacks, and May (1999), inhibition in this context refers to the prevention of automatically-activated but goal-irrelevant information from entering working memory, thus allowing for selective attention while processing. To more directly test this idea, we calculated the number of intrusions in one of our working memory tasks (defined in the context of the task in the Method section) and examined the relationships between working memory intrusions and both MMCB and task performance.

In Gernsbacher's (1990) framework, she asserts that low structure builders' faulty suppression mechanisms are a result of a language-related inhibitory deficit, and cites low structure builders' poor performance on an ambiguity resolution task¹ (compared to high structure builders) as evidence. Recent unpublished research in our lab has examined the

¹ In this task, readers are presented with sentences and are asked to judge whether a target word (presented after each sentence) is related to the meaning of the sentence. In half the trials, the sentence-final word is unambiguous, and the target is unrelated (e.g., "He dug with the shovel"; target = ACE). In the other half of trials, the sentence-final word is ambiguous, and target is related to the contextually inappropriate meaning of that word (e.g., "He dug with the spade"; target = ACE). Gernsbacher et al. (1991) found that low structure builders were slower to reject inappropriate targets (even at longer intervals) relative to high structure builders.

relationship between MMCB and three inhibitory tasks (ambiguity resolution, Flanker, and Stroop; $n = 144$). We found that the MMCB was uncorrelated with the ambiguity resolution ($r = .03$) and Flanker ($r = .03$) tasks, and only moderately correlated with the Stroop ($r = -.22$) task.

While our proxy measures of inhibition may not assess the same construct as the ambiguity resolution task previously used by Gernsbacher, we theorized that they may still play a role in structure building ability and allow for an indirect test of Gernsbacher's (1990) inhibition deficit hypothesis. If low structure builders struggle to suppress irrelevant and unimportant information, then it is possible that individual differences in working memory would explain the variance in structure building ability (and subsequently predict performance on our tasks). Likewise, increased mindwandering, caused by poor inhibitory control, may also pose problems with suppressing irrelevant details that come to mind when building a mental model. We predicted that if Gernsbacher's (1990) hypothesis extends to a general inhibitory deficit (that can be measured using working memory and mindwandering), then these two measures would predict performance on all three tasks. Following the same reasoning, we also believed that the number of working memory intrusions would predict task performance.

Critically, we also manipulated the degree of text structural support offered by the textbook chapter, such that some participants read the chapter in its original form (i.e. high support), while others read a stripped-down plain-text version that contained no bolded terms or end-of-section summary points (i.e. low support). With this manipulation, we aimed to directly test whether the scaffolding currently incorporated into textbooks helps those with lower structure building ability.

We predicted that the degree of text structural support would have interesting implications for low structure builders in the main point identification task. On the one hand, if

Gernsbacher's (1990) account is accurate, and an impoverished ability to suppress irrelevant details accounts for low structure builders' deficits, then perhaps they would identify fewer unimportant points after having read the plain-text chapter, which does not contain as many potential distractors (in the form of end-of-section summaries, for example) as the original-text version. Though it may impact the number of unimportant points identified, we would not predict that text structural support would affect main point identification, if Gernsbacher's (1990) view is correct. If, however, low structure builders have a fundamental problem with identifying main concepts, as we hypothesized, then we would predict that they would fare better when provided high text structural support (in the form of bolded terms and end-of-section summaries) as is the case in the original version of the chapter.

Following Callender and McDaniel (2007) and Bui and McDaniel (2015) – who showed that embedded questions and illustrative diagrams, respectively, helped low structure builders on later inference and problem-solving tasks – we predicted that low structure builders would show especially poor performance on the short-answer test (which requires them to make connections across multiple sections of the chapter) if they had read the plain-text version of the textbook chapter rather than the original version.

Lastly, given Britton and Gulgöz (1991), where the mental models of participants who read a repaired (as opposed to an impaired) text showed stronger alignment with experts' networks, we thought it possible that low structure builders who read the original-text version of the chapter would show greater similarity to experts' ratings in the relatedness ratings task than would those who read the plain-text version. Alternatively, we might not find this text condition by structure building ability interaction for two reasons: firstly, Britton and Gulgöz (1991) did not examine this relationship in the context of structure building ability (and so their findings

may not generalize to the present study), and secondly, neither of our textbook chapters are perfectly analogous to the materials used in their study (i.e. the plain-text chapter contains all the necessary information).

Chapter 2: Method

2.1 Participants and Design

One hundred and thirty-one participants from Washington University in St. Louis took part in the experiment for payment or course credit. To control for differences in baseline knowledge of the material, none of our participants had completed an introductory biology or evolution course (the content of the study material) prior to participating in the experiment. Two participants were excluded from final analyses because they either did not complete the MMCB task (and so we could not ascertain their structure building ability) or did not return for the second session of the experiment.

The degree of structural support offered by the materials was manipulated between participants. Specifically, roughly half the participants ($n = 64$) read an authentic biology textbook chapter in its original format (containing bolded terms and end-of-section main point summaries), while the other half ($n = 65$) read the same chapter in a plain-text format (containing only the text and figures). See Appendix A for a comparison of a page from the original-text chapter and the plain-text chapter.

2.2 Materials

The materials included a textbook chapter, “Evolution and Natural Selection,” taken from the introductory biology textbook, *What is Life? A Guide to Biology* (Phelan, 2009), as used by Wooldridge, Bugg, McDaniel, and Liu (2014).

Structure building ability was assessed using the written portion of the MMCB – consisting of four short fictional narratives, each followed by 12 multiple choice comprehension

questions about important points from the stories. The MMCB takes approximately 20 minutes to complete. Working memory was measured using the automated Operation Span (OSpan) and Reading Span (RSpan) tasks (Unsworth, Heitz, Schrock, & Engle, 2005). In the OSpan, participants were required to solve math problems while trying to remember a series of letters (with set sizes ranging from 3 to 7 math problems/letters long). After each set, participants were asked to recall the letters they had just been presented within that span set, in the correct order. If they could not remember what letter was presented at a certain place in the sequence, they were able to select “blank” as a placeholder. Along with determining accuracy in the OSpan task itself, we also calculated the number of working memory intrusions (defined as any letter selected by participants that was not presented to them during that span set). In the RSpan, participants were also trying to remember a series of letters, but instead of solving math problems in between each letter presentation, they were told to determine if a given word contextually aligns with a presented sentence. Each of these working memory measures takes 15 to 20 minutes to complete. Mindwandering was measured using the Sustained Attention Response Task (SART; Jackson & Balota, 2012), during which participants were rapidly presented with digits from 0 – 9 (in a sequential and random order), and instructed to press the SPACE key any time the presented number was not a 3. Importantly, participants answered 10 on-task/mindwandering probes over the course of the task, from which we calculated mindwandering rates for each participant. The SART takes about 10 minutes to complete.

The short-answer test consisted of 10 integrative deep-level questions, created by the experimenter, assessing understanding and retention of the entire textbook chapter. An example question was “Describe how migration could lead to a change in allele frequencies within a population. How is this different from the founder effect?” (For the full list of questions, please

see Appendix B.) The 10 questions were presented in a random order for each participant. Participants' answers were scored (with each answer earning 0, 1, or 2 points, based on the completeness of responses) using an experimenter-created rubric.

For the main point identification task, several important and unimportant factual concepts from the textbook chapter were identified by the experimenter. An example of an unimportant concept was "The name of Darwin's ship was the *HMS Beagle*" and an example of a main concept was "Fitness depends on an organism's reproductive success compared with other organisms in the population." (For a full list of the 39 concepts, please refer to Appendix C.) Our two experts were asked to independently read through the numbered list, and identify which ones they believed to be the most important concepts. No limits were placed on the time to complete the task or the number of concepts that could be identified. The first expert identified 15 concepts as important, while the second identified 13. There were 10 overlapping concepts that both experts considered important, and nine that only one expert considered important. Participants' identified concepts were compared to experts' to assess their performance on this task.

Lastly, for the relatedness rating task, the experimenter selected 10 main terms from three separate sections of the chapter (the first covering the four mechanisms of evolution, the second covering adaptation to the environment, and the last covering evidence for evolution; please refer to Appendix D for the full list of terms). Both experts were presented with all pairwise combinations of the 30 terms (randomized within three separate blocks, each containing 10 terms), with the order of the pair counterbalanced across experts, and asked to provide relatedness ratings for each pair on a 5-point Likert scale (with 1 meaning *highly unrelated* and 5 meaning *highly related*). Due to a programming error, not all pairwise combinations were

presented to participants (six pairwise combinations were missing in the first set, and three each in the other two sets), and this was taken into account when performing analyses. The Java application program Pathfinder (Schvaneveldt, 1990) was used to determine the proximity of concepts based on relatedness ratings. A proximity matrix was created for each participant and expert, and each matrix was then reflected in a network structure (where each node represents a term; Goldsmith et al., 1991). Separate analyses (assessing similarity between participants' three networks and the respective three average expert networks) were conducted for each set of 10 terms.

2.3 Procedure

The first session consisted of three phases. In the first phase, following the consent process, participants completed the RSpan task. In the second phase, participants completed the SART task. In the last phase, after finishing these cognitive measures, participants were presented with the biology textbook chapter (in PDF form). All participants were instructed to read the chapter once at their own pace, and told that they would be tested on the material two days later, but they were not allowed to take any notes. Participants were given a 15-minute warning prior to the end of their session. Session 1 lasted approximately an hour and a half.

Two days later, participants returned to the laboratory for Session 2, which also consisted of three phases. In the first phase, participants completed the short-answer test, the main point identification task, and the relatedness rating task, in that order. We presented the short-answer test first, as a test of participants' knowledge prior to being presented with any facts from the chapter. We chose to give participants the main point identification task prior to the relatedness ratings task, because the latter provides main terms as part of the procedure and we wanted to test participant's ability to discriminate between unimportant and important concepts beforehand.

Prior to being presented with the short-answer questions, participants were instructed to be succinct in their responses, and to spend no more than 3 to 4 minutes on each question. For the main point identification task, participants were told that the list consisted of facts from the textbook chapter, and their task was to enumerate all the concepts that they believed to be most important. Furthermore, they were told that there was no limit on the number of concepts they could list, and that they did not have to list concepts in any particular order. For the relatedness ratings task, participants were instructed to make their ratings quickly, and to use the full range of the 5-point Likert scale. All three tasks were self-paced.

In the second and third phases of Session 2, participants completed the OSpan and MMCB tasks. In total, Session 2 lasted approximately an hour and a half.

Chapter 3: Results

In our description of the results, we first present multiple regression analyses that examine how well performance on each of our comprehension tasks (main point identification; short-answer conceptual questions; and relatedness ratings) can be predicted by structure building ability (as measured continuously by the MMCB, see Figure 1 for distribution of scores), after accounting for both working memory (as measured by the OSpan and RSpan) and mindwandering (as measured by the SART).

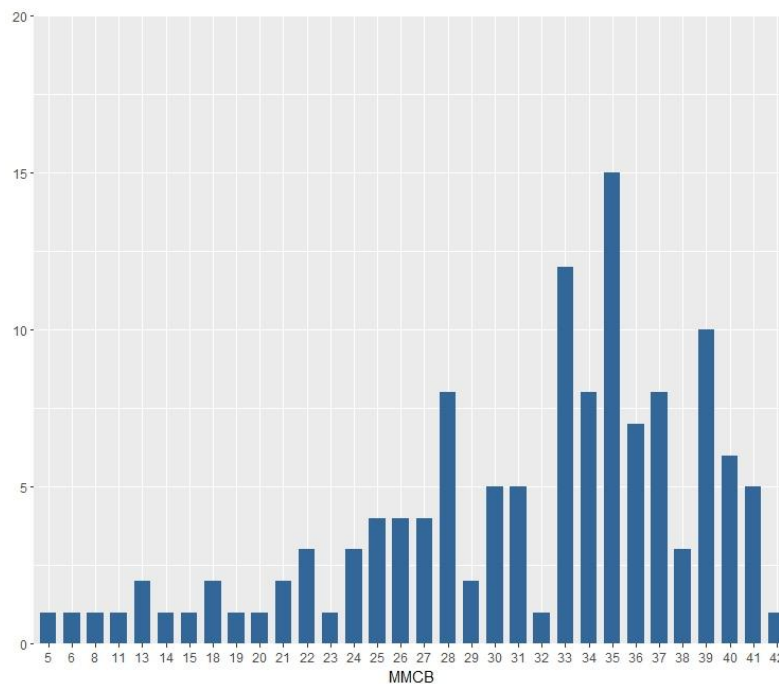


Figure 3.1 Distribution of MMCB scores (ranging from 5 to 42, out of 48).

Importantly, the text condition (our categorical between-subjects factor with two levels: original and plain-text) and the MMCB x text condition interaction term are also included in these analyses. Next, we present multiple regression analyses that determine if our measures of

working memory and mindwandering alone can predict performance on each of our tasks.

Lastly, we present simple regression analyses that examine whether the number of working memory intrusions predict performance on the three tasks.

Following standard protocol, we computed a single working memory score by taking the average of the OSpan and RSpan measures. We then tested the relationship between this composite working memory score and MMCB. As shown in Figure 2, we replicated prior research, and found a modest but significant correlation ($r = .20$, $t(121) = 2.23$, $p = .02$) between the two measures.

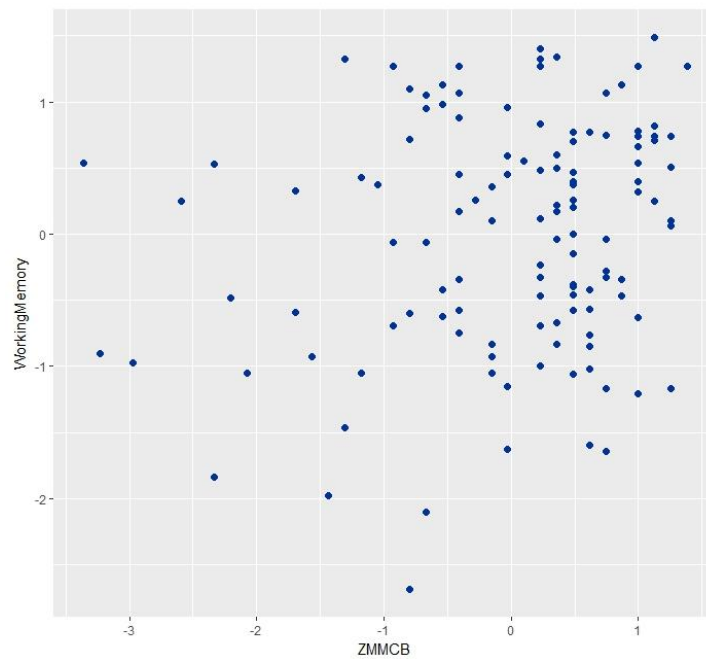


Figure 3.2 Relationship between standardized MMCB and working memory. There was a small but significant positive correlation ($r = .20$) between the two measures.

We also tested the relationship between SART scores and MMCB, and as shown in Figure 3, found there to be a marginally significant negative correlation ($r = -.15$, $t(126) = -1.66$, $p = .10$).

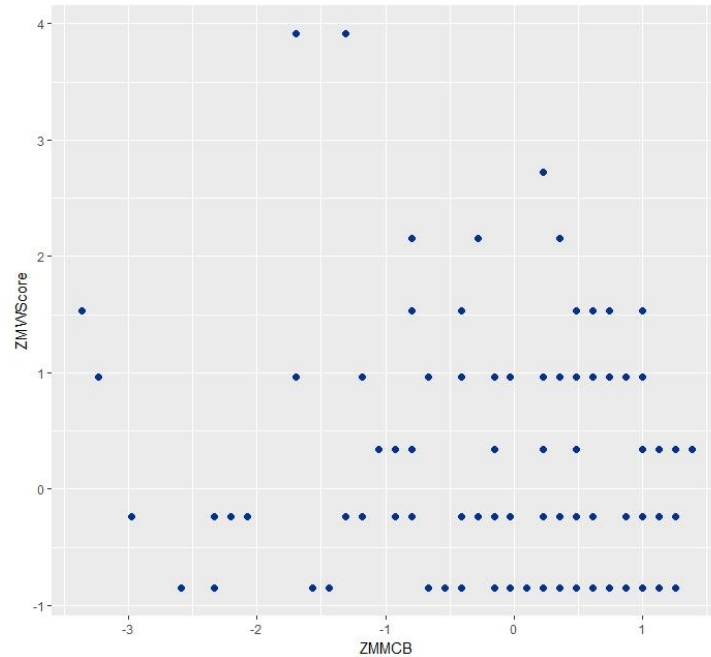


Figure 3.3 Relationship between standardized MMCB and mindwandering. There was a marginally significant negative correlation ($r = -.15$) between the two measures.

Lastly, we found a small and nonsignificant correlation between OSpan intrusions and MMCB ($r = -.12$, $t(122) = -1.30$, $p > .05$), as shown in Figure 4.

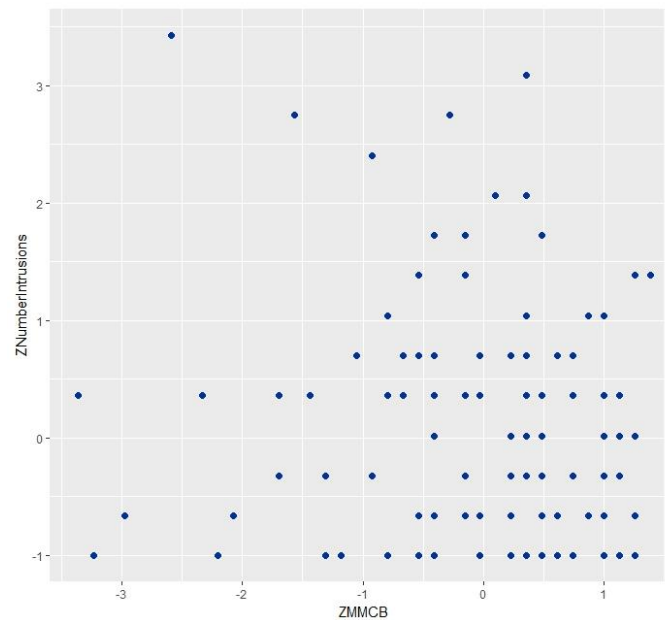


Figure 3.4 Relationship between standardized MMCB and number of working memory intrusions. There was a nonsignificant negative correlation ($r = -.12$) between the two measures.

We calculated descriptive statistics of the critical measures (unstandardized; see Table 1), and then computed simple Pearson correlations of our predictors and dependent variables (see Table 2) before conducting the regressions described above.

Table 3.1 Descriptive Statistics of Predictors and Dependent Variables in Tasks

	Descriptive Statistics		
	<i>M</i>	<i>SD</i>	<i>Range</i>
MMCB	31.21	7.80	37
OSpan	53.40	16.11	75
RSpan	47.14	16.95	68
Mindwandering	.18	.21	1
Main Point Id Total	13.08	7.52	37
Main Point Id Hits	7.24	2.14	10
Main Point Id Partial Hits	3.74	2.45	9
Main Point Id False Alarms	2.52	4.43	23
Short Answer Score	15.20	3.85	17
Relatedness Ratings Task 1 Similarity	.40	.12	.67
Relatedness Ratings Task 2 Similarity	.26	.08	.51
Relatedness Ratings Task 3 Similarity	.30	.11	.63

Table 3.1 Note that mindwandering and similarity scores are proportions (ranging from 0 to 1). The max possible scores for all the other variables is as follows: MMCB = 48; OSpan = 75; RSpan = 75; Main Point Id Total = 39; Main Point Id Hits = 10; Main Point Id Partial Hits = 9; Main Point Id False Alarms = 20; Short Answer Score = 20.

Table 3.2 Correlations of Predictors and Dependent Variables in All Tasks

	MMCB	Mindwandering	OSpan	RSpan	OSpan Intrusions	False Alarms	d' (Hits)	Short Answer Score	Similarity Set 1	Similarity Set 2	Similarity Set 3
MMCB											
Mindwandering	-0.147										
OSpan	0.236**	-0.001									
RSpan	0.125	-0.007	0.533***								
OSpan Intrusions	-0.117	0.045	0.140	0.197*							
False Alarms	-0.085	0.081	-0.063	-0.087	-0.031						
d' (Hits)	0.291***	-0.134	0.120	0.068	-0.028	-0.599***					
Short Answer	0.368***	-0.117	0.115	0.082	0.074	-0.020	0.169				
Score											
Similarity Set 1	0.380***	-0.188*	0.177*	0.122	-0.155	-0.197*	0.238**	0.205*			
Similarity Set 2	0.102	0.062	0.010	-0.137	-0.063	-0.040	0.116	0.157	0.079		
Similarity Set 3	0.319***	-0.088	0.334***	0.193*	-0.045	-0.158	0.267**	0.238**	0.367***	0.229**	

Computed correlation used Pearson-method with pairwise-deletion.

Table 3.2 Significant findings ($p < .05 = *$; $p < .01 = **$; $p < .001 = ***$) denoted in black. Nonsignificant findings denoted in grey. Note that working memory measures here are reported separately, and not as a computed average score.

All variables were standardized for the purpose of analyses, and the level of significance was set at $\alpha = .05$. Regression tables and graphs were created using the sjPlot package in the software program *R* (Lüdtke, 2017).

3.1 Main Point Identification Task

In the following analyses, adapting from terminology used in signal detection literature, if a participant identified one of the 10 concepts that both biology experts considered important, this is denoted a “hit.” If a participant identified one of the nine concepts that only one biology expert considered important, this is called a “partial hit.” And lastly, if the participant identified a concept that neither expert deemed important, this is termed a “false alarm.” Following signal detection literature, we calculated d' , a commonly used sensitivity index that measures the difference between the standardized hit rate and standardized false alarm rate (Macmillan & Creelman, 2004) for each participant. Analyses of d' values calculated using hits versus d' values calculated using both hits and partial hits are reported separately.

First, we determined if structure building ability could predict participants' hits, after accounting for working memory and mindwandering. We conducted a multiple regression (with MNCB, working memory, mindwandering, text condition, and the MNCB x text condition interaction term as predictors), and found that MNCB was a significant predictor, $\beta = .32$, $t(112) = 2.42$, $p = .02$. Therefore, at average levels of working memory and mindwandering, for every one standard deviation increase in MNCB, there was a .32 standard deviation increase in number of hits. We also found a significant main effect of text condition, $\beta = -.35$, $t(112) = -1.96$, $p = .05$, such that, at average levels of working memory, mindwandering, and MNCB, participants in the plain-text condition identified .35 standard deviations fewer hits than those in the original-text condition. There were no main effects of working memory or mindwandering, and the MNCB x

text condition interaction was not significant, both $ps > .05$. As shown in Table 3, this model accounted for 7.6% of the variance (adjusted $R^2 = .076$).

	Number of Hits		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.17	-0.08 – 0.42	.170
Working Memory	-0.01	-0.22 – 0.20	.951
Mindwandering	-0.04	-0.22 – 0.14	.650
Text Condition	-0.35	-0.71 – 0.00	.052
MMCB	0.31	0.06 – 0.56	.017
Text Condition:MMCB	-0.03	-0.39 – 0.33	.883
Observations	118		
R^2 / adj. R^2	.116 / .077		

Table 3.3 Results showing the main effects of text condition and MMCB on the standardized number of hits in the main point identification task, accounting for working memory and mindwandering. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

We next conducted a multiple regression predicting participants' hits with just working memory and mindwandering as our predictors, and found that neither cognitive measure was a significant predictor, both $ps > .05$. Lastly, we conducted a simple regression to determine if participants' working memory intrusions predicted hits, and found that they did not, $p > .05$.

We then conducted a multiple regression to determine if structure building ability predicted participants' false alarms, after accounting for all the other measures, and found that, as shown in Table 4, there were no significant main effects or interactions, all $ps > .05$.

Table 3.4 Null Effects of Predictors on False Alarms			
	Number of False Alarms		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.06	-0.20 – 0.32	.653
Working Memory	-0.10	-0.32 – 0.13	.394
Mindwandering	0.06	-0.13 – 0.25	.510
Text Condition	-0.07	-0.45 – 0.31	.721
MMCB	0.01	-0.26 – 0.28	.925
Text Condition:MMCB	-0.14	-0.53 – 0.24	.459
Observations	118		
R^2 / adj. R^2	.021 / -.023		

Table 3.4 Results showing the null effects of all predictors on the standardized number of false alarms in the main point identification task.

We then determined if working memory and mindwandering predicted false alarms, and once again, found that neither measure was a significant predictor, both $ps > .05$. The same was true of working memory intrusions, $p > .05$.

Next, we conducted a multiple regression to determine if MMCB could predict d' values (calculated using hits, which were concepts identified by both experts), after accounting for all other predictors, and found that MMCB was a marginally significant predictor, $\beta = .30$, $t(121) = 1.94$, $p = .06$. These results indicate that, at average levels of working memory and mindwandering, for every one standard deviation increase on the MMCB, there was a .30 standard deviation increase in d' . There were no main effects of working memory, mindwandering, or text condition, nor was the MMCB x text condition interaction significant, both $ps > .05$. As shown in Table 5, the overall model accounted for 8.3% of the variance (adjusted $R^2 = .083$).

Table 3.5 Effect of MMCB on d'			
	d' (Hits)		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.11	-0.18 – 0.41	.453
Working Memory	0.09	-0.16 – 0.34	.480
Mindwandering	-0.10	-0.32 – 0.11	.335
Text Condition	-0.28	-0.71 – 0.14	.192
MMCB	0.30	-0.01 – 0.60	.055
Text Condition:MMCB	0.12	-0.32 – 0.55	.593
Observations	118		
R ² / adj. R²	.122 / .083		

Table 3.5 Results showing the main effect of MMCB on d' in the main point identification task, accounting for working memory, mindwandering, and text condition. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R²), are noted in bold.

We then conducted a multiple regression to test whether working memory or mindwandering could account for individual differences in d', and found that there were no significant main effects of either factor, both $ps > .05$. Lastly, we determined if OSpan intrusions could predict d', and found that they did not, $p > .05$.

Next, we recalculated d' (subtracting the standardized false alarm rate from the sum of standardized hit and standardized partial hit rates) for a more liberal assessment of main point identification, and re-ran the same regression analyses as above. We found that MMCB significantly predicted d', $\beta = .45$, $t(116) = 2.25$, $p = .03$. At average levels of working memory and mindwandering, for a one standard deviation increase in MMCB, there was a .45 standard deviation increase in this more liberally-calculated d'. There was also a marginally significant main effect of text condition, $\beta = -.55$, $t(116) = -1.96$, $p = .05$, such that, at average levels of

working memory, mindwandering, and MMCB, participants in the plain-text condition showed a .55 standard deviation decrease in d' relative to those in the original text condition. There were no main effects of working memory or mindwandering, and the MMCB x text condition interaction was not significant, all $ps > .05$. As shown in Table 6, this model accounted for 8.4% of the variance (adjusted $R^2 = .084$). In our next regression analysis, we found that neither working memory nor mindwandering significantly predicted d' values, both $ps > .05$. And in our last regression analysis, we found that working memory intrusions also did not predict these liberal d' values, $p > .05$.

Table 3.6 Effects of Text Condition and MMCB on Liberal d'

	d' (Hits & Partial Hits)		
	B	CI	p
(Intercept)	0.25	-0.14 – 0.64	.208
Working Memory	0.07	-0.26 – 0.40	.676
Mindwandering	-0.01	-0.30 – 0.27	.919
Text Condition	-0.55	-1.10 – 0.01	.052
MMCB	0.45	0.05 – 0.84	.026
Text Condition:MMCB	0.08	-0.48 – 0.65	.768
Observations	122		
R^2 / adj. R^2	.122 / .084		

Table 3.6 Results showing the main effects of text condition and MMCB on a more liberally calculated d' in the main point identification task, accounting for working memory, mindwandering, and text condition. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

Taken together, these results indicate that those with lower structure building ability struggle (relative to high structure builders) to identify main concepts (as identified by both or only one of the experts), even after taking working memory and mindwandering into account. Further, their impaired performance on this task cannot be accounted for by inhibition deficits as indexed by working memory, mindwandering, or working memory intrusions. Moreover, receiving a lower degree of text structural support from the textbook chapter also led to decreased performance on the task, after accounting for working memory, mindwandering, and structure building ability.

3.2 Short-Answer Test

We developed a rubric to score participants' answers on the short-answer test, where each of the 10 answers was assigned a score of 0, 1, or 2. Two research assistants independently scored five randomly selected participants' tests (blind to both the structure building ability and text condition of the participants) according to the rubric. Inter-rater reliability was high ($r = .88$), and so they each scored half of the remaining participants' tests.

To start, we conducted a multiple regression to determine if MMCB could predict participants' standardized short-answer test performance (after accounting for all other predictors), and found, contrary to our predictions, no significant main effects of MMCB, text condition, working memory, or mindwandering, all $ps > .05$. Critically, there was a significant MMCB x text condition interaction, $\beta = 1.56$, $t(115) = 2.36$, $p = .02$, such that (as shown in Figure 5), those with lower structure building ability performed especially poorly compared to those with higher structure building ability in the plain-text condition relative to the original-text condition.

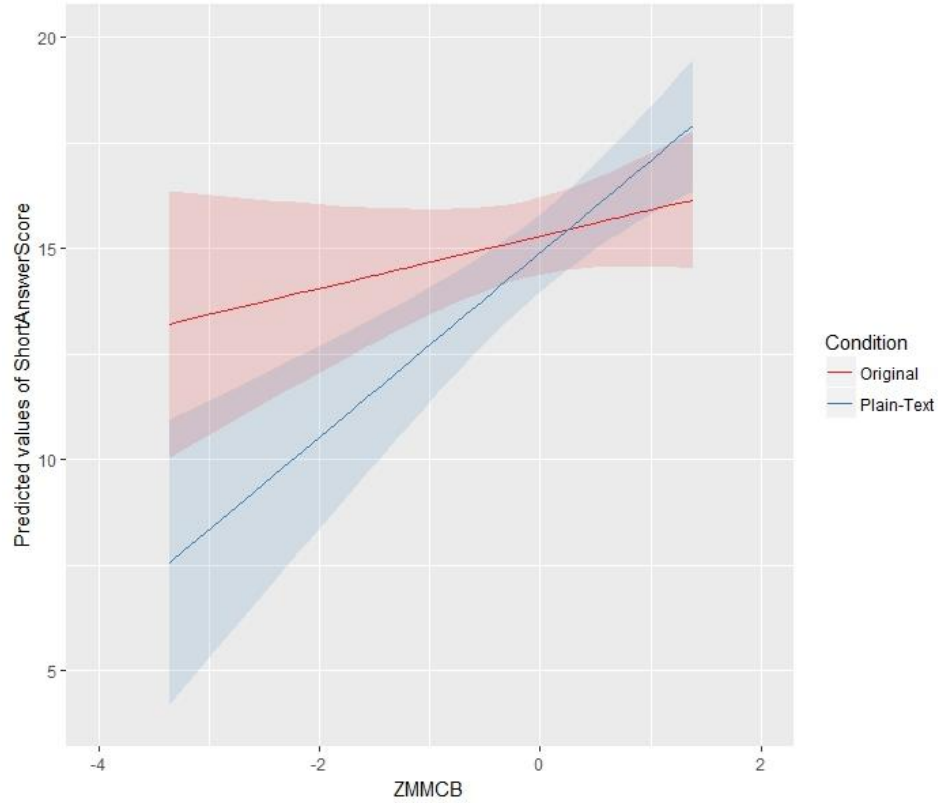


Figure 3.5 Interaction between text condition and standardized MMCB on short-answer test performance, accounting for working memory and mindwandering.

Overall, as shown in Table 7, our model accounted for 15.1% of the variance (adjusted $R^2 = .151$). After sub-setting the data by condition, we found, after accounting for working memory and mindwandering, a marginally significant effect of MMCB in the original-text condition, $\beta = .85$, $t(57) = 1.79$, $p = .07$, and a significant effect of MMCB in the plain-text condition, $\beta = 2.20$, $t(56) = 4.59$, $p < .001$.

Table 3.7 Interaction of Text Condition and MMCB on Short Answer Test Performance

	Short Answer Score		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	15.29	14.38 – 16.21	<.001
Working Memory	0.34	-0.42 – 1.11	.377
Mindwandering	-0.27	-0.93 – 0.39	.415
Text Condition	-0.39	-1.69 – 0.90	.550
MMCB	0.62	-0.30 – 1.54	.185
Text Condition:MMCB	1.56	0.25 – 2.88	.020
Observations	121		
R² / adj. R²	.186 / .150		

Table 3.7 Results showing the interaction between text condition and MMCB on short-answer test performance, accounting for working memory, mindwandering, and text condition. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R²), are noted in bold.

Together, these results indicate that, when provided a low degree of text structural support, low structure builders are less able to integrate knowledge across different parts of their mental models relative to high structure builders; increased text support, however, mitigates some differences on task performance. We then conducted a multiple regression to examine if working memory and mindwandering could predict participants' short-answer test performance, and found that, once again, neither measure was a significant predictor, both $ps > .05$. Lastly, we found that working memory intrusions also did not predict short-answer test performance, $p > .05$.

Lastly, to test whether individual differences in short-answer test scores could be entirely explained by participants' ability to identify main points, we conducted a multiple regression with d' (using only hits, as decided a priori because these were items that both experts

considered important), MMCB, and the MMCB x text condition interaction term as our predictors. There were no significant main effects of main point identification accuracy (as measured by d') or text condition, both $ps > .05$. MMCB, however, was a marginally significant predictor, $\beta = .71$, $t(119) = 1.65$, $p = .10$. At average levels of main point identification accuracy, for every one standard deviation increase in MMCB, there was a .71 standard deviation increase in short-answer test performance. Importantly, as shown in Figure 6, the MMCB x text condition interaction also remained significant, $\beta = 1.43$, $t(119) = 2.34$, $p = .02$, such that those with lower structure building ability performed especially poorly relative to those with higher structure building ability on the short-answer test in the plain-text condition relative to those in the original-text condition.

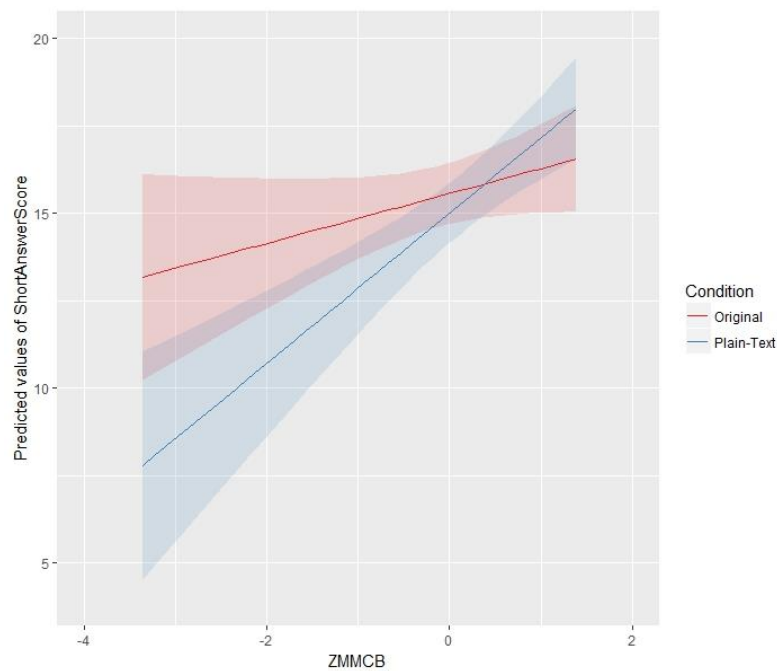


Figure 3.6 The interaction between text condition (original-text, plain-text) and standardized MMCB on short-answer test performance, accounting for main point identification accuracy (d'). The interaction shows that, even after taking performance on the main point identification task into consideration, low structure builders who read the plain-text version of the chapter performed worse on the short-answer test.

Overall, as shown in Table 8, this model accounted for 16.6% of the variance (adjusted $R^2 = .166$).

Table 3.8 Interaction of Text Condition and MMCB on Short-Answer Test Performance, Accounting for d'

	Short Answer Score (by d' and MMCB)		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	15.57	14.70 – 16.45	<.001
Text Condition	-0.55	-1.78 – 0.67	.374
MMCB	0.71	-0.14 – 1.57	.101
d' (Hits)	0.16	-0.37 – 0.70	.553
Text Condition:MMCB	1.44	0.22 – 2.65	.021
Observations	124		
R² / adj. R²	.193 / .166		

Table 3.8 Results showing the main effect of MMCB and the interaction between text condition and MMCB on short-answer test performance, accounting for main point identification accuracy (d') and text condition. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

3.3 Relatedness Ratings Task

Before we present the analyses, we first describe how we accounted for the previously mentioned programming error. Not all pairwise combinations were presented to participants in all three blocks, and because Pathfinder cannot accommodate missing data, we needed to impute the data for all missing pairwise combinations. We first calculated participants' average rating across all pairs they were provided, and found this to be 3 in all three sets. Before replacing all missing values with 3s, however, we used the experts' complete rating sets to determine how much of an impact this may have. We compared the coherence of experts' original models, and compared these to their networks' coherence values after having replaced the missing-for-

participants data points with 1s, 5s, and 3s (i.e., the two extreme ratings and the average rating). In all three sets, replacing the experts' original data points with 3s had the least impact on network coherence (see Table 9 for calculations using the first set). Therefore, in the following analyses, we replaced all missing participant data with 3s, but used the experts' original ratings when deriving an average median expert network.

Table 3.9 Network Coherence with Imputed Data in First Relatedness Rating Set

	Expert 1	Expert 2
Original	0.23	0.63
Replaced with 1s	0.23	0.06
Replaced with 5s	0.09	0.59
Replaced with 3s	0.24	0.60

Table 3.9 Note that replacing original values with 3s had the least overall impact.

We first validated experts' networks by calculating individual network coherence and path link correlations for each of the three relatedness rating sets. The coherence values of experts' networks were greater than .20 (which is the necessary minimum value) in the first and third relatedness ratings sets, but not the second (as shown in Table 10).

Table 3.10 Coherence of Experts' Networks in All Three Relatedness Ratings Tasks

	Expert 1	Expert 2
Relatedness Rating Set 1	0.23	0.63
Relatedness Rating Set 2	0.009	0.20
Relatedness Rating Set 3	0.40	0.34

Table 3.10 Note that coherence values in the second relatedness ratings set did not meet accepted standard requirements (need to be above .20).

The path link correlations, which measure the overlap between each individual expert network and the average median expert network, were greater than the necessary minimum value of .50 (which is required in order to include any individual expert network in average median calculations) for all three relatedness ratings sets (as shown in Table 11). In general, these

validation measures are necessary to ensure that no one expert unduly influences the average median network, which is used as a comparison to participants' networks (Neiles, Todd, & Bunce, 2016).

Table 3.11 Path-link Correlations of Experts' Networks to Average Median Network in All Three Relatedness Ratings Tasks

	Expert 1	Expert 2
Relatedness Rating Set 1	0.80	0.91
Relatedness Rating Set 2	0.81	0.69
Relatedness Rating Set 3	0.88	0.89

Table 3.11 Note that path-link correlations in all three relatedness ratings task sets were above the commonly accepted standard of .50.

We calculated similarity scores (between participants' networks and the average median expert for each set) for all three relatedness ratings sets. (See Figure 7 for the average median expert network of the first relatedness ratings set, as well as an example each of lowly and highly similar participant networks.)

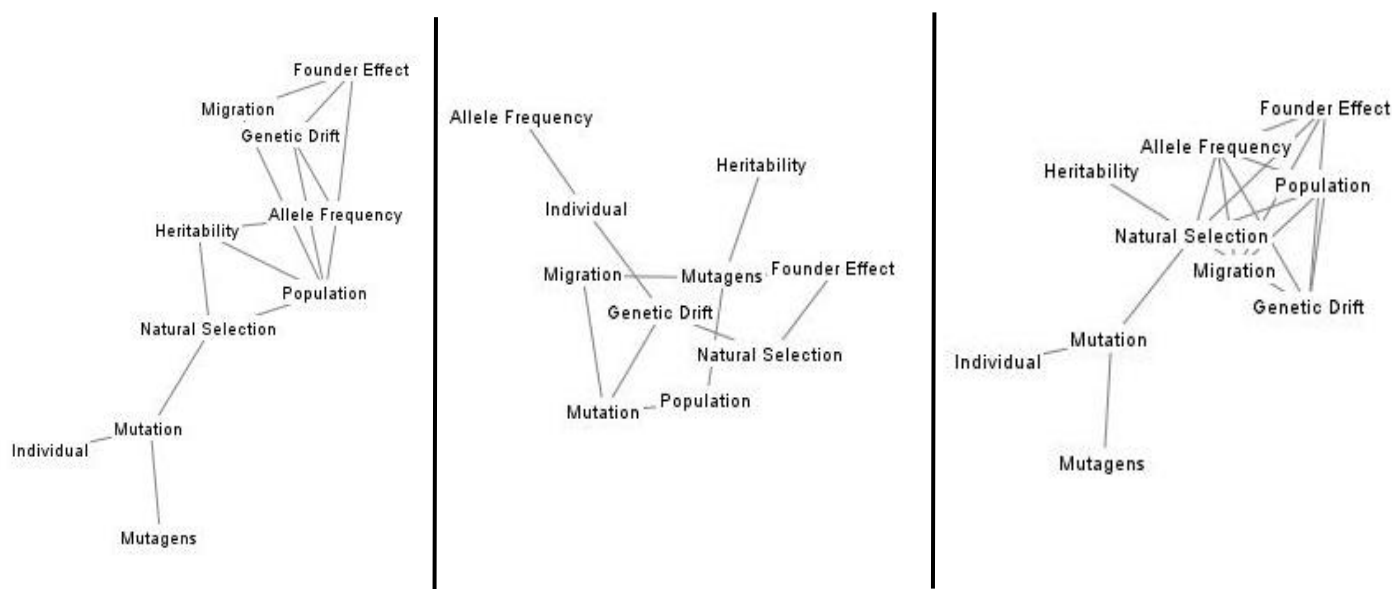


Figure 3.7. The average median expert network for the first relatedness ratings task set (left panel); an example of a lowly similar participant network (middle; standardized similarity score = -3.48); and an example of a highly similar participant network (right; standardized similarity score = 2.28).

We conducted a multiple regression to determine if structure building ability would predict similarity scores on the first relatedness ratings set, after accounting for working memory and mindwandering (with text condition and a MMCB x text condition interaction term also included in the model). MMCB was a significant predictor, $\beta = .24$, $t(116) = 2.08$, $p = .04$, showing that, at average levels of working memory and mindwandering, for a one standard deviation increase in MMCB, there was a .24 standard deviation increase in similarity scores. There were no significant main effects of text condition, working memory, or mindwandering, or a significant MMCB x text condition interaction, all $ps > .05$. As shown in Table 12, this model accounted for 13.3% of the variance (adjusted $R^2 = .133$).

Table 3.12. Effect of MMCB on Similarity Scores in Relatedness Rating Task Set 1

	Similarity Scores – Relatedness Ratings Set 1		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.01	-0.21 – 0.24	.918
Working Memory	0.13	-0.06 – 0.32	.173
Mindwandering	-0.12	-0.28 – 0.05	.158
Text Condition	0.06	-0.26 – 0.38	.705
MMCB	0.24	0.01 – 0.47	.038
Text Condition:MMCB	0.15	-0.18 – 0.47	.368
Observations	122		
R^2 / adj. R^2	.169 / .133		

Table 3.12. Results showing the main effect of MMCB on standardized similarity scores in the first relatedness ratings task set, accounting for working memory, mindwandering, and text condition. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

We then conducted a multiple regression to determine if working memory and mindwandering also predicted similarity scores on the first relatedness ratings set. There was a

marginally significant main effect of working memory, $\beta = .19$, $t(119) = 1.93$, $p = .06$, such that, at average levels of mindwandering, a one standard deviation increase in working memory yielded a .19 standard deviation increase in similarity scores. There was also a marginally significant main effect of mindwandering, $\beta = -.16$, $t(119) = -1.91$, $p = .06$, such that, at average levels of working memory, a one standard deviation increase in mindwandering resulted in a .16 standard deviation decrease in similarity scores. As shown in Table 13, this model accounted for 4.3% of the variance in similarity scores (adjusted $R^2 = .043$).

Table 3.13. Effects of Working Memory and Mindwandering on Similarity Scores in Relatedness Ratings Task Set 1

	Similarity Scores – Relatedness Ratings Set 1		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.04	-0.12 – 0.21	.602
Working Memory	0.19	-0.01 – 0.38	.057
Mindwandering	-0.16	-0.33 – 0.01	.058
Observations	122		
R^2 / adj. R^2	.059 / .043		

Table 3.13. Results showing the effects of working memory and mindwandering on standardized similarity scores in the first relatedness ratings set. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

We then conducted a simple regression analysis to determine if working memory intrusions predicted similarity scores in this set. We found a marginally significant main effect of intrusions, $\beta = -.15$, $t(119) = -1.73$, $p = .09$, such that a one standard deviation decrease in intrusions resulted in a .15 standard deviation decrease in similarity scores.

We conducted the same three regression analyses on the second relatedness ratings set, and found that there were no significant main effects of MNCB, working memory, mindwandering, or OSpan intrusions. There was a marginally significant effect of text condition,

$\beta = .32$, $t(116) = 1.77$, $p = .08$, such that, at average levels of working memory, mindwandering, and MMCB, participants in the plain-text condition showed a .32 standardized unit increase in similarity scores than those in the original-text condition. As shown in Table 14, this model accounted for 1.2% of the variance (adjusted $R^2 = .012$). Because of the low coherence of the experts' individual networks, we caution against drawing strong conclusions from these results.

Table 3.14 Effect of Text Condition on Similarity Scores in Relatedness Rating Set 2

	Similarity Scores – Relatedness Ratings Set 2		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	-0.11	-0.36 – 0.14	.365
Working Memory	-0.12	-0.33 – 0.09	.258
Mindwandering	0.08	-0.10 – 0.27	.359
Text Condition	0.32	-0.04 – 0.68	.079
MMCB	0.16	-0.09 – 0.41	.213
Text Condition:MMCB	-0.08	-0.45 – 0.28	.648
Observations	122		
R² / adj. R²	.053 / .012		

Table 3.14 Results showing the main effect of text condition on standardized similarity scores in the second relatedness ratings task set, accounting for working memory, mindwandering, and MMCB. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

We once again conducted a multiple regression to determine if structure building ability predicted similarity scores in the third relatedness ratings set (after accounting for all other predictors). We found that MMCB was a significant predictor, $\beta = .34$, $t(116) = 2.83$, $p = .006$, such that, at average levels of working memory and mindwandering, a one standard deviation increase in MMCB resulted in a .34 standard deviation increase in similarity scores. We found that working memory was also a significant predictor, $\beta = .27$, $t(116) = 2.79$, $p = .006$, such that,

at average levels of mindwandering and MMCB, a one standard deviation increase in working memory resulted in a .27 standard deviation increase in similarity scores. There were no significant main effects of text condition or mindwandering, or a significant MMCB x text condition interaction, all $ps > .05$. As shown in Table 15, this model accounted for 14.7% of the variance (adjusted $R^2 = .147$).

Table 3.15 Effects of Working Memory and MMCB on Similarity Scores in Relatedness Rating Set 3

	Similarity Scores – Relatedness Ratings Set 3		
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.15	-0.08 – 0.38	.198
Working Memory	0.27	0.08 – 0.47	.006
Mindwandering	-0.04	-0.21 – 0.13	.611
Text Condition	-0.22	-0.56 – 0.11	.181
MMCB	0.34	0.10 – 0.57	.006
Text Condition:MMCB	-0.16	-0.50 – 0.17	.337
Observations	122		
R² / adj. R²	.182 / .147		

Table 3.15 Results showing the effects of working memory and MMCB on standardized similarity scores in the third relatedness ratings task set, accounting for mindwandering and text condition. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

We then conducted a multiple regression to determine if working memory and mindwandering predicted similarity scores, and found that there was a significant main effect of working memory, $\beta = .35$, $t(119) = 3.56$, $p < .001$, such that, at average levels of mindwandering, a one standard deviation increase in working memory yielded a .35 standard deviation increase in similarity scores. There was no significant main effect of mindwandering, $p > .05$. As shown

in Table 16, this model accounted for 8.8% of the variance in similarity scores (adjusted $R^2 = .088$). Lastly, we found that working memory intrusions did not predict similarity scores in the third relatedness ratings set, $p > .05$.

Table 3.16 Effects of Working Memory on Similarity Scores in Relatedness Ratings Set 3

Similarity Scores – Relatedness Ratings Set 3			
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.03	-0.14 – 0.20	.697
Working Memory	0.35	0.15 – 0.54	<.001
Mindwandering	-0.08	-0.26 – 0.09	.332
Observations	122		
R^2 / adj. R^2	.103 / .088		

Table 3.16 Results showing the effects of working memory on standardized similarity scores in the third relatedness ratings task set, accounting for mindwandering. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

Finally, to test whether MMCB remained a significant predictor after accounting for participants' ability to identify main points in the main point identification task, we conducted multiple regressions, with d' (calculated using hits), MMCB, text condition, and the MMCB x text condition interaction term as our predictors, on similarity scores in all three relatedness ratings sets. We found that MMCB significantly predicted similarity scores in the first set (as shown in Table 17), $\beta = .33$, $t(120) = 2.88$, $p = .004$, and in the third set (as shown in Table 18), $\beta = .37$, $t(120) = 3.17$, $p = .002$, but not in the second (as shown in Table 19), $p > .05$. Moreover, we found that d' was only a significant predictor in the third relatedness ratings set, $\beta = .16$, $t(120) = 2.11$, $p = .04$, but not in the first or second.

Table 3.17 Effect of MNCB on Similarity Scores in Relatedness Rating Set 1, Accounting for d'

Similarity Scores – Relatedness Ratings Set 1 (by d' and MNCB)			
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	-0.09	-0.32 – 0.14	.457
Text Condition	0.11	-0.22 – 0.43	.523
MNCB	0.33	0.10 – 0.56	.005
d' (Hits)	0.11	-0.03 – 0.26	.119
Text Condition:MNCB	0.07	-0.25 – 0.39	.665
Observations	125		
R^2 / adj. R^2	.189 / .162		

Table 3.17 Results showing the main effect of MNCB on standardized similarity scores in the first relatedness ratings task set, accounting for main point identification accuracy (d'). Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

Table 3.18 Effect of Text Condition on Similarity Scores in Relatedness Rating Set 2, Accounting for d'

Similarity Scores – Relatedness Ratings Set 2 (by d' and MNCB)			
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	-0.13	-0.37 – 0.12	.317
Text Condition	0.30	-0.05 – 0.65	.092
MNCB	0.04	-0.20 – 0.29	.741
d' (Hits)	0.10	-0.05 – 0.25	.206
Text Condition:MNCB	-0.01	-0.36 – 0.33	.940
Observations	125		
R^2 / adj. R^2	.039 / .007		

Table 3.18 Results showing the main effect of text condition on standardized similarity scores in the second relatedness ratings task set, accounting for MNCB and main point identification accuracy (d'). Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

Table 3.19 Effects of MMCB and d' on Similarity Scores in Relatedness Rating Set 3

Similarity Scores – Relatedness Ratings Set 3 (by d' and MMCB)			
	<i>B</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.09	-0.15 – 0.32	.468
Text Condition	-0.14	-0.48 – 0.19	.396
MMCB	0.37	0.14 – 0.61	.002
d' (Hits)	0.16	0.01 – 0.30	.037
Text Condition:MMCB	-0.26	-0.59 – 0.07	.122
Observations	125		
R^2 / adj. R^2	.149 / .120		

Table 3.19 Results showing the main effects of MMCB and main point identification accuracy (d') on standardized similarity scores in the third relatedness ratings task set, accounting for text condition. Marginally significant and significant effects, as well as proportion of variance explained (adjusted R^2), are noted in bold.

Lastly, we found that there was a marginally significant main effect of text condition only in the second relatedness ratings set, $\beta = .30$, $t(120) = 1.70$, $p = .09$, such that, at average levels of main point identification accuracy and MMCB, participants in the plain-text condition showed a .30 standard deviation increase in similarity scores relative to those in the original text condition. The MMCB x text condition interaction was not significant in any of the relatedness ratings sets, $p > .05$.

Taken together, these results indicate that, even when given important terms from the text (i.e., nodes in a structure), those with lower structure building ability struggle to build expert-like structures around these terms. While working memory appears to account for some of the variance in similarity scores, it is clear that an inhibitory deficit (as indexed by working memory, mindwandering, and working memory intrusions) cannot explain the whole story. Furthermore,

participants' performance on the main point identification task does not fully account for variance in their ability to build networks similar to experts' when provided main terms.

Chapter 4: Discussion

In the present study, we were interested in identifying whether low structure builders' deficits are only related to faulty suppression mechanisms (as asserted by Gernsbacher, 1990) or if they experience deficits in the first step of the structure building process. We used an educationally authentic textbook chapter, and designed three tasks (main point identification, short-answer test, and relatedness ratings) that specifically focused on different parts of the structure building process, as outlined in Gernsbacher (1990)'s framework. Moreover, we manipulated the degree of text structural support in the textbook chapter, to offer insight into potential interventions that may help bolster low structure builders' abilities. We believed that those with low structure building ability would show impaired performance on all three tasks relative to those with high structure building ability, and that these differences may be particularly pronounced for low structure builders who read the plain-text version of the textbook chapter, as opposed to the original version. Critically, we tested if structure building ability would predict task performance after accounting for working memory capacity and mindwandering. Lastly, we used these cognitive measures as proxies for inhibition to indirectly test if all performance deficits could be explained by Gernsbacher's (1990) inhibitory deficit hypothesis.

4.1 Main Point Identification Task

We found that structure building ability predicted the number of participants' hits and the more liberal d' (calculated using hits and partial hits), and was marginally predictive of the more conservative d' (calculated using only hits), even after accounting for working memory and mindwandering. Structure building ability did not significantly predict the number of false

alarms. Importantly, we found that working memory, mindwandering, or OSpan intrusions did not predict the number of participants' hits or false alarms, or either of the d' measures.

These findings directly counter the predictions of Gernsbacher's (1990) account of structure building, which asserts that low structure building ability can be fully explained by a faulty suppression mechanism, for several reasons. Firstly, if low structure builders struggle because they are unable to suppress unimportant concepts as they build a mental model, we would have expected structure building ability and/or our proxy inhibition measures to predict participants' number of false alarms (because those are concepts that neither expert considered to be important), but none of the measures did. To reiterate, low structure builders did not select a significantly higher number of unimportant points than high structure builders, which poses serious problems for Gernsbacher's view.

Secondly, if all individual differences in structure building ability can be explained by an imperfect suppression mechanism, then there should have been no structure-building-related performance differences in participants' ability to identify the main points of the textbook chapter. Moreover, to the extent that working memory and mindwandering are indices of inhibition, the findings further indicate that an inhibitory deficit cannot explain performance differences on this task.

There are several parts of the structure building process that could be implicated in main point identification. According to Gernsbacher's (1990) structure building framework, the initial process of laying a solid foundation involves the activation of important memory nodes, which are tied to topic sentences or content words. Moreover, the enhancement mechanism (which allows for the continued activation of important memory nodes, through the use of anaphoric devices that highlight the importance of some concepts over others) could also play a role in

identifying main points. In the context of the present study, perhaps low structure builders are less adept than high structure builders at comprehending topic sentences, content words, or anaphora. Future research could directly test this possibility, by assessing low structure builders' memory for and understanding of these key devices that are fundamental to the activation of important memory nodes.

Another possible factor that may influence main point identification is prior knowledge, which may help alert readers to the importance of certain concepts over others. Some prior research has also connected prior knowledge to the suppression of irrelevant information: McNamara and McDaniel (2004) showed that readers with greater domain-specific or general knowledge showed less interference on an ambiguity resolution task, regardless of reading ability, than readers with less prior knowledge. We did not attempt to replicate these findings in the present study, but we found structure-building related differences in task performance even after controlling for prior knowledge in the context of formalized college-level training in biology and evolution. It is also possible that low structure builders struggle to *activate* what prior knowledge they do have relative to high structure builders, which would then negatively impact their ability to activate foundational nodes. Indeed, recent unpublished data in the lab ($n = 144$) assessing the relationship between MMCB and the Remote Associates Test (which has been implicated in knowledge activation; Wiley, 1998) revealed a significant correlation ($r = .49, p < .001$) between the two measures.

Interestingly, we found that text condition significantly predicted the number of hits (concepts both experts identified as important) and marginally significantly predicted the liberal d' (which was calculated using both hits and partial hits, concepts that only one expert labeled important), after accounting for working memory, mindwandering, and structure building ability.

Specifically, participants who read the plain-text version of the chapter identified fewer important points and performed worse on the task. Notably, all hits and partial hits contained terms that were either bolded or reiterated in pre-chapter or post-section summaries, which were all forms of text structural support not offered in the plain-text chapter. Therefore, it appears that the structural scaffolding currently existent in textbooks does indeed help alert both low and high structure builders to the key points of a chapter.

4.2 Short-Answer Test

We predicted that low structure builders would show impaired performance on the short-answer test, for several reasons. They might struggle to answer deep-level conceptual questions because of our theorized main point identification deficit (which would result in impoverished models) or because, as Gernsbacher (1990) asserts, of faulty suppression mechanisms (which would result in cluttered models). In both cases, we would expect that low structure builders would be less able to effectively use their mental model (be it impoverished or cluttered) and answer integrative questions than high structure builders. Surprisingly, we found no significant main effect of structure building; instead we found a significant MMCB x text condition interaction, which revealed that low structure builders fared especially poorly (relative to high structure builders) on the short-answer test after having read the plain-text chapter than the original chapter. Notably, when we subset the data by condition, we found a marginally significant effect of MMCB on test performance in the original-text condition and a significant effect of MMCB in the plain-text condition. These findings indicate that text structural support devices help readers (especially low structure builders) answer deep-level integrative questions about the chapter.

We then determined if structure building would predict short-answer test performance after accounting for participants' ability to identify main points (using the conservative d' values

from the first task), for two reasons. Firstly, we were interested to see if a main point identification deficit could fully account for individual differences in short-answer test performance, as we hypothesized. Secondly, we theorized (based on findings from the main point identification task) that text scaffolding helps to highlight important concepts, by bolding important terms and including them in end-of-section summaries. If structure building predicted short-answer test score after accounting for performance on the main point identification task, then this would reveal that the main point identification impairment does not capture all of low structure builders' deficits, and moreover, that text scaffolding aids in more than the highlighting of main points. Indeed, we found MNCB to be a marginally significant predictor of short-answer test performance in this case. Again, we believe these findings reveal that the ability to answer conceptual short-answer questions requires more than simply being able to identify the main points, and that the text structural support devices aid in these processes as well. Importantly, the MNCB x text condition interaction remained significant after taking main point identification into consideration, underscoring the importance of text structural support.

We found that neither working memory, mindwandering, nor working memory intrusions predicted participants' short-answer test performance, which potentially contradicts Gernsbacher's (1990) view that an inhibitory deficit can entirely explain low structure building ability. According to Gernsbacher, low structure builders would perform worse on a short-answer test requiring the integration of knowledge from different sections of the textbook chapter, because of their poor suppression mechanisms resulting in too-cluttered models. If this were the case, then we would have expected these proxy measures of inhibition to predict performance on the short-answer test. To truly disconfirm this hypothesis, however, further

research using either an ambiguity resolution task or more traditional measures of inhibition would need to be conducted.

4.3 Relatedness Ratings Task

We found structure building ability to be a significant predictor of similarity scores in both the first and third relatedness ratings set, even after accounting for both working memory and mindwandering. These findings indicate that, when referring back to mental models they formed while reading the textbook chapter in order to make relational judgments between important terms in the text, low structure builders struggle to make expert-like relatedness ratings. In other words, the connections (between foundational nodes) in their mental models are not as similar to the experts' as connections in high structure builders' representations. We found working memory and mindwandering to be marginally significant predictors of similarity scores in the first relatedness ratings set, and found working memory to be a significant predictor of similarity in the third relatedness ratings set, after accounting for both mindwandering and MMCB. These findings indicate that low performance on our proxy measures of inhibition contributed to an impaired ability to make the same relational judgments as experts, but does not fully account for all variance in similarity scores.

We were again interested in determining if performance on the relatedness ratings task (where participants are provided all the important terms) could be entirely accounted for by performance on the main point identification task (where participants are required to discriminate between unimportant and important concepts), and found that they could not. In other words, structure building ability significantly predicted similarity scores in the first and third relatedness ratings sets, even after accounting for main point identification accuracy (measured by the conservative d'). These findings indicate that, as was the case with the short-answer test, the

ability to identify the main points (i.e., lay the initial foundation) is not the sole factor contributing to the ability to make connections between these foundational nodes in an expert-like manner.

4.4 What do Individual Differences in Structure Building Involve?

The results of the present study overwhelmingly confirm that structure building entails multiple skills. Gernsbacher (1990) outlines three structure building processes: laying a foundation of important points; mapping incoming information onto developing structures; and, in cases where incoming information does not cohere, shifting to create new structures. She also highlights two mechanisms (enhancement and suppression) that play a role in increasing and decreasing, respectively, the activation of a foundational memory node. Structure building, then, not only involves identifying important concepts, but also discriminating between important and unimportant concepts and making connections across foundational nodes while expanding upon a mental model. The findings of the present study align with this basic framework, and underscore the importance of both identifying the main points and integrating across them. While we have a basic understanding of what factors influence participants' decisions about whether to map incoming information onto an existing structure or shift to form new structures (Anderson et al., 1983; Gernsbacher & Robertson, 1994), further research is necessary to better understand what skills are involved in making connections across nodes within an overall mental model.

Because of the many processes involved in successful structure building, there are multiple points at which low structure builders can struggle. Although Gernsbacher (1990) claims that all individual differences in structure building can be attributed to a faulty suppression mechanism, the results of the present study clearly contradict such an account of

structure building. Indeed, the findings of the present study provide new insight into what characterizes a low structure builder. We determined that those with low structure building ability falter at the very first step of the structure building process, and experience a main point identification deficit, which has, to date, never been empirically shown. Furthermore, because we found that this main point identification deficit does not predict individual differences in performance on our other two tasks, we believe that low structure builders also struggle with knowledge integration. These findings align with previous research showing that low structure builders perform worse on inference multiple-choice and problem-solving questions, as well as on free recall questions, because presumably all of these tasks required the integration of knowledge across parts of readers' mental models (Bui & McDaniel, 2015; Martin et al., 2016). Because we showed that structure building ability predicted performance on our tasks even after accounting for working memory and mindwandering, we strongly suggest that low structure builders do not have general cognitive impairments, and that their deficits are specifically tailored to the skills involved in structure building. While, as Gernsbacher (1990; 1994) suggests, low structure builders may also struggle with suppressing irrelevant and unimportant information (though we found no evidence for this in the present study), identifying an additional deficit that characterizes low structure building helps bring us closer to elucidating underlying causal mechanisms.

4.5 Educational Implications of Textbook Scaffolding

We manipulated the degree of text structural support in the textbook chapter to determine if the scaffolding currently existent in textbooks – bolded words, pre-chapter summary, and end-of-section summaries, for example – helps bring low structure builders' performance up to the level of high structure builders. The present study revealed several important findings that shed light

on how this scaffolding already helps readers, and, critically, how it can be improved to specifically bolster low structure builders' abilities.

We found that participants performed worse on the main-point identification task after having read the plain-text chapter than did those who read the original chapter, regardless of structure building ability. In other words, it seems as though incorporating bolded words and end-of-section summaries into the text truly helps alert readers to the important concepts, which has major educational implications. Because the impact of structure building was not modulated by text condition, however, it is clear that more scaffolding is necessary to equate low and high structure builders' performance on this task. Given the findings of the study, we believe that low structure builders would benefit from even more explicit highlighting of important points throughout the chapter. We also suggest that these practices could be incorporated into the classroom as well, because it may be possible that low structure builders struggle to identify main points during lectures.

On the short-answer test, we found a significant interaction between structure building ability and text structure, such that low structure builders performed especially poorly relative to high structure builders after reading the plain-text chapter. Again, it appears as though being exposed to bolded terms and end-of-section summaries helped readers make connections across multiple sections of the chapter, leading to better performance on a test that required a deep conceptual understanding of the text as a whole. Because structure building, surprisingly, did not predict performance on the test, it is possible that the degree of text structural support present in the original chapter was enough to bolster low structure builders' performance to the level of high structure builders. However, given past findings of performance differences on comprehension and memory tests between low and high structure builders when given

educationally authentic materials (Callender & McDaniel, 2007), we believe that textbooks could incorporate more scaffolding to help facilitate the process of making connections across the text.

Lastly, we did not find a reliable main effect of text condition on our relatedness ratings task, but structure building predicted similarity scores, meaning that current textbook practices need further improvement. The relatedness ratings task forces readers to refer back to the mental model built while reading the textbook chapter, and then make relational judgments about important concepts that occur within the same section. Because low structure builders made less similar judgments to experts than did high structure builders regardless of what text condition they were in, we believe that textbooks should work to help readers better understand how concepts are related. Drawing from the findings of Bui and McDaniel (2015), it is possible that incorporating illustrative diagrams that demonstrate the relations between concepts into textbook chapters may assist low structure builders with this task.

Taken together, these findings strongly suggest that scaffolding helps alert readers to the main points of the text and make connections between different sections, but, critically, the degree of text structural support in the original textbook chapter is still not enough to allow low structure builders to fully overcome their deficits.

4.6 Conclusion

We designed three tasks (main point identification; short-answer test; relatedness ratings) aimed at identifying potential deficits experienced by low structure builders (relative to high structure builders) at different parts of the structure building process. In her structure building framework, Gernsbacher (1990) asserts that all individual differences in structure building arise from low structure builders having faulty suppression mechanisms. However, in the present study, we determined that low structure builders falter, in fact, at an even earlier part of the process: they

struggle to identify the overarching main points (which serve as foundational nodes for their structures) of a text, relative to high structure builders.

We predicted that this main point identification deficit would have negative cascading effects on later aspects of structure building, which involve making connections across main points, but our findings indicate that this deficit cannot account for all individual differences in structure building. More specifically, participants' performance on the main point identification task did not predict performance on the short-answer test or on two (out of three) relatedness ratings tasks. Clearly, good structure building ability requires multiple skills, and a single deficit – be it faulty suppression, as Gernsbacher (1990) claims, or main point identification impairment, as hypothesized in the present study – cannot tell the whole story. If we hope to fully determine where in the structure building process low structure builders falter, then further research is necessary to explore and tease apart these, and other, potential deficits.

We also indirectly tested Gernsbacher's (1990) account that individual differences in structure building arise entirely due to low structure builders' inhibitory deficits. Participants' working memory and mindwandering served as proxy measures of inhibition in our study, and we tested if these measures could predict performance on any of our comprehension measures. We also tested if participants' working memory intrusions predicted task performance, for the same reasons. Because these measures do not directly assess language-related inhibitory mechanisms (unlike, for example, the ambiguity resolution task; Gernsbacher, 1990), we cannot draw firm conclusions negating Gernsbacher's hypothesis. However, given the findings of the present study, as well as unpublished data from our lab that showed small correlations between more traditional measures of inhibition (Stroop and Flanker tasks) and MMCB, we strongly suggest that a general inhibitory deficit does not underlie low structure building.

Notably, we also found that the degree of text structural support significantly affected participants' performance on our tasks, such that, in most cases, having bolded terms and end-of-section summary points helped readers identify main points and draw connections across multiple sections. These findings affirm current textbook practices, and indicate that manipulating the amount of scaffolding provided in the text (or perhaps even in other classroom materials) in future studies is a beneficial avenue to determine how to mitigate some differences between low and high structure builders.

By identifying that low structure builders experience a deficit in the earliest part of the structure building process – and that this deficit alone cannot explain performance differences on tasks related to later parts of the structure building process – we are able to better understand how low structure builders differ in their abilities from high structure builders. However, further research is necessary to identify other potential deficits, and even more importantly, causal mechanisms underlying structure building. By focusing on causal mechanisms, we will be able to design better interventions in the future that can specifically target low structure builders' deficits and bolster their abilities.

References

1. Agarwal, P. K., Finley, J. R., Rose, N.S., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, 6, 764-771.
2. Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology*, 35A, 427-440.
3. Arnold, K. M., Daniel, D. B., Jensen, J. L., McDaniel, M. A., & Marsh, E. J. (2016). Structure building predicts grades in college psychology and biology. *Applied Cognitive Psychology*, 30, 454-459.
4. Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 18, 467-484.
5. Britton, B. K., & Gulgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83, 329-345.
6. Bui, D. C., & McDaniel, M. A. (2015). Enhancing learning during lecture note-taking using outlines and illustrative diagrams. *Journal of Applied Research in Memory and Cognition*, 4, 129-135.
7. Callender, A. A., & McDaniel, M. A. (2007). The benefits of embedded question adjuncts for low and high structure builders. *Journal of Educational Psychology*, 99, 339-348.
8. Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, 3, 422-433.
9. Foss, D. J., & Lynch, R. H., Jr. (1969). Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception and Psychophysics*, 5, 145-148.
10. Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
11. Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 430-445.
12. Gernsbacher, M. A. & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 245-262.
13. Gernsbacher, M. A., & Robertson, R. W. (1994). The definite article "the" facilitates mapping in story comprehension. *Bulletin of the Psychonomic Society*, 29, 496

14. Gernsbacher, M. A., & Varner, K. R. (1988). *The multi-media comprehension battery*. Eugene, OR: University of Oregon, Institute of Cognitive and Decision Sciences.
15. Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83, 88-96.
16. Haberlandt, K. (1984). Components of sentence and word reading times. In D.E. Kieras & M.A. Just (Eds.), *New methods in reading comprehension research* (pp. 219-252). Hillsdale, NJ: Erlbaum.
17. Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G.H. Bower (Vol. Ed.), *The psychology of learning and motivation* (pp. 193-225). New York, NY: Academic Press.
18. Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriati (Vol. Eds.), *Attention and performance: XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 653-675). Cambridge, MA: MIT Press.
19. Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in younger and older adults: Converging evidence from the sustained attention to response task and reading for comprehension. *Psychology & Aging*, 27, 106-119.
20. Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169-183.
21. Lüdtke, D. (2017). sjPlot: Data visualization for statistics in social science. *R package version 2.4.0*.
22. Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Psychology Press.
23. Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin and Review*, 1, 126-129.
24. Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of web-based and lecture courses. *Journal of Experimental Psychology: Applied*, 8, 85-98.
25. Martin, N. D., Nguyen, K., & McDaniel, M. A. (2016). Structure building differences influence learning from educational text: Effects on encoding, retention, and metacognitive control. *Contemporary Educational Psychology*, 46, 52-60.
26. McDaniel, M. A., Hines, R. J., and Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language*, 46, 544-561.

27. McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 465-482.
28. McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive control. *Journal of Experimental Psychology: General*, 141, 302-320.
29. Neiles, K. Y., Todd, I., & Bunce, D. M. (2016). Establishing the validity of using network analysis software for measuring students' mental storage of chemistry concepts. *Journal of Chemical Education*, 93, 821-831.
30. Phelan, J. (2009). *What is life? A guide to biology*. New York, NY: W. H. Freeman.
31. Redick, T. S., Heitz, R. P., & Engle, R. W. (2007). Working memory capacity and inhibition: Cognitive and social consequences. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 125-142). Washington, DC: American Psychological Association.
32. Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex Publishing.
33. Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505.
34. Wiley, J. (1988). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition*, 26, 716-730.
35. Wooldridge, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Journal of Applied Research in Memory and Cognition*, 3, 214-221.

Appendix A

Text Comparison (with original-text on left and plain-text on right)

8.14

Artificial selection is a special case of natural selection.

In practice, plant and animal breeders understood natural selection before Darwin, they just didn't know that they understood it. Farmers bred crops for maximum yield, and dog, horse, and pigeon fanciers selectively bred the animals with their favorite traits to produce more and more of the offspring with more and more exaggerated versions of the trait.

The artificial selection used by animal breeders and farmers is also natural selection because the three conditions are satisfied, even though the differential reproductive success is being determined by humans and not by nature. Apple growers, for example, use artificial selection to produce the wide variety of apples available: green, yellow, and red, tart and sweet, large and small. What is important is that it is still differential

reproductive success, and the results are no different. It was a stroke of genius for Darwin to recognize that the same process farmers were using to develop new and better crop varieties was occurring naturally in every population on earth and always had been.

TAKE-HOME MESSAGE 8.14

Animal breeders and farmers are making use of natural selection when they modify their animals and crops, because the three conditions for natural selection are satisfied. Since the differential reproductive success is determined by humans and not by nature, this type of natural selection is also called artificial selection.

In practice, plant and animal breeders understood natural selection before Darwin, they just didn't know that they understood it. Farmers bred crops for maximum yield, and dog, horse, and pigeon fanciers selectively bred the animals with their favorite traits to produce more and more of the offspring with more and more exaggerated versions of the trait.

The artificial selection used by animal breeders and farmers is also natural selection because the three conditions are satisfied, even though the differential reproductive success is being determined by humans and not by nature. Apple growers, for example, use artificial selection to produce the wide variety of apples available: green, yellow, and red, tart and sweet, large and small. What is important is that it is still differential reproductive success, and the results are no different. It was a stroke of genius for Darwin to recognize that the same process farmers were using to develop new and better crop varieties was occurring naturally in every population on earth and always had been.

8.15

Certain traits are easily categorized—some people have blue eyes and others have brown eyes, just as some tigers have white fur and some have orange fur. Other traits, such as height in humans, are influenced by many genes and environmental factors so that a continuous range of phenotypes occurs. Whether a given trait is influenced by one gene or by a complex interaction of many genes and the environment, it can be subject to natural selection and be changed in any of several ways.

In directional selection, individuals with one extreme from the range of variation in the population have higher fitness. Milk production in cows is an example. There is a lot of variation in milk production from cow to cow. As you might expect, farmers select for breeding those cows with the highest milk production and have done so for many decades. The result of such selection is not surprising: between 1920 and 1945, average milk production increased by about 50% in the United States.

8.15

Natural selection can change the traits in a population in several ways.

Certain traits are easily categorized—some people have blue eyes and others have brown eyes, just as some tigers have white fur and some have orange fur. Other traits, such as height in humans, are influenced by many genes and environmental factors so that a continuous range of phenotypes occurs (FIGURE 8-24). Whether a given trait is influenced by one gene or by a complex interaction of many genes and the environment, it can be subject to natural selection and be changed in any of several ways.

Directional Selection In directional selection, individuals with one extreme from the range of variation in the population have higher fitness. Milk production in cows is an example. There is a lot of variation in milk production from cow to cow. As you might expect, farmers select for breeding those cows with the highest milk production and have done so for many decades. The result of such selection is not surprising: between 1920 and 1945, average milk production increased by about 50% in the United States (FIGURE 8-25).

FIGURE 8-24 Some traits fall into clear categories, others range continuously.

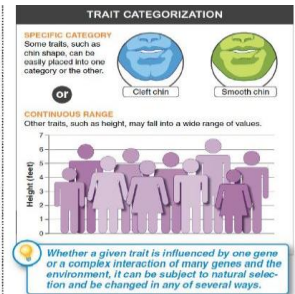
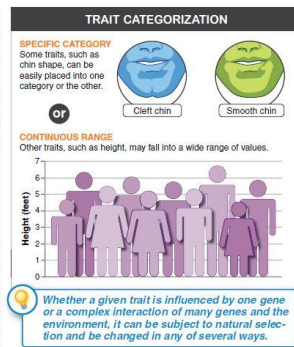


FIGURE 8-24 Some traits fall into clear categories, others range continuously.

Appendix B

Short-Answer Test Questions

1. How does the variation in finches' beaks help explain biogeography?
2. Why do some organisms evolve characteristics through natural selection that don't seem useful or functional (for example, some flies don't use 1% of their wings)?
3. The cheetahs of today are almost lacking in genetic variation. How is this possible? What mechanism of evolution is likely responsible according to the chapter you read?
4. Why is penicillin becoming less effective as an antibiotic?
5. Describe how migration could lead to a change in allele frequencies within a population. How is this different from the founder effect?
6. In general, are mutations helpful or harmful for living things? Why?
7. Can evolution be demonstrated in a lab? Use an example from the book to explain.
8. Explain why evolution occurs at the population level and not at the individual level. Mention allele frequencies in your explanation.
9. If a wolf has a trait that allows it to survive for twice as long as other wolves, but it cannot reproduce, does this wolf have a high 'fitness'? Why or why not?
10. What was the significance of the Galápagos finches to Darwin's theory of evolution?

Appendix C

Main Point Identification Task Concepts

1. Fruit flies can be bred to live a long time without food
2. Charles Lyell wrote *Principles of Geology*
3. Variation for a trait, heritability, and differential reproductive success are necessary for natural selection
4. Genetic drift has greatest impact in small populations
5. Alfred Wallace also posited evolution by natural selection
6. Adaptation increases fitness
7. Cheetahs suffered a population bottleneck 10,000 years ago
8. Some bacteria have evolved to be resistant to antibiotics
9. Random mutations are the ultimate source of genetic variation in a population
10. Humans can cause evolution through artificial selection
11. Polydactyly-Amish individuals more frequently have extra fingers and toes
12. After his trip to the Galapagos, Charles Darwin wrote *On the Origin of Species*
13. Migration (or gene flow) is the movement of some individuals of a species from one population to another
14. Glyptodont fossils resemble the armadillos of today
15. Fitness depends on an organism's reproductive success compared with other organisms in the population
16. The average beak size of Galapagos finches is always changing
17. Directive selection and disruptive selection are two forms of natural selection where certain individuals in a population experience the highest fitness
18. One to 2% of a fly's wings do not help it to fly
19. Ultraviolet radiation can cause DNA mutations
20. Biogeography describes the patterns in the geographic distribution of living organisms
21. The earth is about 4.6 billion years old
22. Homologous structures sometimes end up having little or no function at all
23. There are roughly 125 amino acid differences between humans and lamprey eels
24. Natural selection does not lead to perfect organisms

25. The fossil record offers evidence for natural selection by helping to identify “missing links” between groups of species
26. Turkeys on poultry farms can no longer mate naturally
27. There is almost no genetic variation left in the current population of cheetahs
28. The founder effect occurs when the founding members of a new population can have different allele frequencies than the original source population
29. Mutations are very rare
30. Evolution is a change in allele frequencies in a population
31. The evolutionary history of horses is very well preserved
32. Jean Baptiste Lamarck also suggested species change over time
33. The name of Darwin’s ship was the *HMS Beagle*
34. Before Darwin, most people believed that all species had been created separately and were unchanging
35. Comparative anatomy and embryology reveal common evolutionary origins
36. Herbert Spencer first devised the term “survival of the fittest”
37. The issue of whether radiation from cell phones is harmful is still being debated
38. Natural landmarks can influence migration
39. Adaptation is the process by which organisms become better matched to the environment, as well as the specific features that make those organisms more fit

Appendix D

List of Terms in Relatedness Ratings Tasks

Set 1

1. individual
2. population
3. allele frequency
4. mutation
5. genetic drift
6. migration
7. natural selection
8. mutagens
9. founder effect
10. heritability

Set 2

1. fitness
2. genotype
3. phenotype
4. reproductive output
5. adaptation
6. directional selection
7. stabilizing
8. selection
9. disruptive selection
10. differential reproductive success
11. selective pressure

Set 3

1. biogeography
2. embryology
3. radiometric dating
4. fossil
5. homologous structure
6. vestigial structure
7. convergent evolution
8. evolutionary clock
9. comparative anatomy
10. evolutionary family tree